U.PORTO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
UNIVERSIDADE DO PORTO

MHC in sea bass: molecular and structural insights to class I antigen presentation

Rute Daniela Pinheiro da Silva Pinto

Tese de doutoramento em Ciências Biomédicas

2013

Rute Daniela Pinheiro da Silva Pinto

MHC in sea bass: molecular and structural insights to class I antigen presentation

Tese de Candidatura ao grau de Doutor em Ciências Biomédicas submetida ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

Orientador – Doutor Nuno Miguel Simões dos Santos

Categoria – Investigador Auxiliar

Afiliação – Instituto de Biologia Molecular e Celular

Co-orientador – Doutor Pedro José Barbosa Pereira

Categoria – Investigador Auxiliar

Afiliação – Instituto de Biologia Molecular e Celular

De acordo com o disposto no Decreto-Lei nº 74/2006 de 24 de Março, esclarece-se ser da nossa responsabilidade a execução das experiências que estiveram na origem dos resultados apresentados, assim como a sua interpretação, discussão e redacção.
Nesta tese foram utilizados os resultados dos artigos publicados ou submetidos para publicação abaixo indicados.

1. **Pinto RD**, Randelli E, Buonocore F, Pereira PJ, Dos Santos NM. (2013) Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) MHC class I heavy chain and β2-microglobulin. Dev Comp Immunol. 39 (3): 234-254, doi:10.1016/j. dci.2012.10.002.
2. **Pinto RD**, Pereira PJ, dos Santos NM. (2011) Transporters associated with antigen processing (TAP) in sea bass (*Dicentrarchus labrax*, L.): molecular cloning and characterization of TAP1 and TAP2. Dev Comp Immunol. 35:1173-81.
3. **Pinto RD**, da Silva DV, Pereira PJ, dos Santos NM. (2012) Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) Tapasin. Fish Shellfish Immunol. 32 (1):110-20.
4. **Pinto RD**, Moreira AR, Pereira PJ, dos Santos NM. (2013) Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) calreticulin. Fish Shellfish Immunol. *in press* doi: 10.1016/j.fsi.2013.03.004.
5. **Pinto RD**, Moreira AR, Pereira PJ, dos Santos NM. (2013) Two thioredoxin family members from sea bass (*Dicentrarchus labrax*, L.): molecular characterization of PDI (PDIA1) and ERp57 (PDIA3). Fish Shellfish Immunol. *submitted*.

Acknowledgements

This work has been carried out in the Fish Immunology and Vaccinology Group at IBMC-INEB Associate Laboratory. During this time I have been funded by FCT PhD fellowship BD/42327/2007, financed by POPH-QREN and co-funded by FSE and MCTES, which I acknowledge.

I wish to thank Nuno dos Santos and Pedro Pereira for their involvement with this project in the first place and for guidance.

I am thankful to colleagues, co-authors and other collaborators that contributed to this work and/or supported me in some way, making this thesis possible.

I want to mention present and former colleagues at the FIV Lab for the challenges faced and time spent together throughout years: Ana, Daniela, Marisa, Raquel, Leonor, Marta, Carolina, Márcio and Eunice.

I leave a special word to Raquel also for answering my dull requests while I was writing. The help of Eunice, Frederico Silva and Catarina Almeida was crucial for the execution of the preliminary experiments detailed in the Appendix section.

To my brothers, sisters-in-law, niece and nephews I thank all the crazy and joyful (Satur)days.

Finally, I am truly grateful to my parents for their unconditional and most valuable support. I dedicate this work to them.

Rute

Contents

Abbreviations list

Ab	antibody
ABC	ATP-binding cassette
Ag	antigen
AgR	antigen receptor
AID	activation-induced cytidine deaminase
AMP	antimicrobial peptide
APC	antigen presenting cell
ATP	adenosine tri-phosphate
α1	antigen binging region of the MHC class I heavy chain
α2	antigen binging region of the MHC class I heavy chain
α3	β2-microglobulin binging region of the MHC class I heavy chain
BCR	B cell receptor
BLAST	basic local alignment search tool
bm	bone marrow
bp	base pair
β2m	β2-microglobulin
CALR	calreticulin
CDA	cytidine deaminase
cDNA	complementary DNA
CD	cluster of differentiation
CH/CL	coupling helices or loops
CLIP	class II-associated invariant chain peptide
CLP	common lymphoid progenitor
CLR	C-type lectin receptor
CMP	common myeloid progenitor
CNX	calnexin
CP	connective peptide
CRT	calreticulin
CTL	cytotoxic T lymphocyte
CYT	cytoplasmic
DC	dendritic cell
DEPC	diethylpyrocarbonate
Dila	*Dicentrarchus labrax*
DNA	deoxyribonucleic acid

DRiP	defective ribosomal products
Dscam	Down syndrome cell adhesion protein
ER	endoplasmic reticulum
ERAAP	ER aminopeptidase associated with antigen processing
ERAD	ER-associated degradation
ERAP	ER aminopeptidase
ERp57	ER protein of 57 kDa
EST	expressed sequence tag
FAO	Food and Agriculture Organisation
FcRn	neonatal Fc receptor
FREP	fibrinogen-related protein
GALT	gut-associated lymphoid tissues
gDNA	genomic DNA
ght	general hematopoietic tissues
HC	heavy chain
HK	head kidney
HLA	human leukocyte antigen
HSC	hematopoietic stem cell
Ia	classical class I molecules
Ib	non-classical class I molecules
ICL	intracytoplasmic loop
ICOS	inducible costimulator
IFN	interferon
Ig	immunoglobulin
IgC	Ig constant domain
IgSf	Ig superfamily
IgV	Ig variable domain
Ii	invariant chain
indel	insertion/deletion
kbp	kilobase pairs
kDa	kiloDalton
LMP	low molecular mass protein (proteasome)
LRR	leucine-rich repeat
MALT	mucosal-associated tissues
MH	major histocompatibility
MHC	major histocompatibility complex
MHC-II	MHC class II

MIIC	MHC-II containing compartment
mRNA	messenger RNA
Mya	million years ago
NBD	nucleotide binding domain
NFκB	nuclear factor κB
NHEJ	non-homologous end joining
NK	natural killer
NLR	NOD-like receptor
NLS	nuclear localization signal
ORF	open reading frame
PAMP	pathogen-associated molecular pattern
PBD	peptide binding domain
PBL	peripheral blood lymphocytes
PBR	peptide binding region
PCR	polymerase chain reaction
PDB	protein data bank
PDI	protein disulfide isomerase
pI/Mw	isoelectric point/molecular weight
PLC	peptide loading complex
poly(A)	polyadenylation
polyI:C	polyinosinic-polycytidylic acid
PRR	pattern recognition receptor
RACE	rapid amplification of cDNA ends
RAG	recombination activation gene
RFLP	restriction fragment length polymorphism
RLR	RIG-I-like receptor
RNA	ribonucleic acid
RSS	recombination signal sequence
RT-PCR	reverse transcribed-PCR
SD	standard deviation
SR	sarcoplasmic reticulum
SRCR	scavenger receptor cysteine rich
SSE	secondary structure elements
TAP	transporter associated with antigen processing
TAPBP	TAP binding protein
TEC	thymic epithelial cell
TCR	T cell receptor

TdT	terminal deoxynucleotidyl transferase
T_H	T helper cells
TLR	Toll-like receptor
TM	transmembrane
TMD	transmembrane domain
TPN	tapasin
TPN-R	tapasin-related protein
Trx	thioredoxin
UTR	untranslated region
VCBP	variable region-containing chitin-binding proteins
V(D)J	variable (diversity) joining
VLR	variable lymphocyte receptor
3D	three-dimensional

Summary

All organisms rely on immune defences in order to protect themselves from a variety of pathogens. However, the tools and mechanisms available to accomplish an immune response vary among animal life. Depending on the species level of organization, several mechanisms can be simultaneously employed. Vertebrates arose approximately 500 Mya, being presently divided in jawless (lampreys and hagfish) and jawed vertebrates (cartilaginous and bony fish, amphibians, reptiles, birds and mammals). The hallmark of this group of organisms is the emergence of a new kind of immune cell - the lymphocyte - bearing somatically diversified and clonally expressed antigen receptors, in which each lymphocyte is armed with a single receptor. As a consequence, a vast repertoire of Ag receptors/immune cells can be generated. Two distinct lineages of lymphocytes are present, B and T cells, with functions in humoral (antibodies) and cellular immunity, respectively. In the case of jawed vertebrates, the Ag receptors are based on immunoglobulin (Ig) domains. B cell receptors (BCRs) exist at the surface of B cells, while T cell receptors (TCRs) are present on the plasma membrane of T cells. T cells only recognize antigens in the context of major histocompatibility molecules (MHC). These are divided in MHC class I (MHC-I) and class II (MHC-II) molecules generally serving to present endogenous peptides to CD8+ T cells and exogenous peptides to CD4+ T lymphocytes, respectively. Nearly all nucleated cells have class I complexes, while only professional antigen presenting cells (APCs) contain the class II ones. Dendritic cells, the most important type of APC, are indeed a jawed vertebrate innovation. Besides this important function of sampling peptides, reporting the state of a cell, MHC is also crucial for the education of T lymphocytes, a process that takes place in the thymus, a novel vertebrate-specific immune-related organ.

MHC class I molecules are formed by the heavy chain (HC), β2-microglobulin and a loaded short peptide. Assembly is a tightly regulated process that occurs within the endoplasmic reticulum (ER) with assistance of common chaperones and dedicated factors. Early oxidative folding of the HC involves calnexin and ERp57. The subsequently formed HC/β2m heterodimer is highly unstable in the absence of peptides. Calreticulin (CRT), ERp57, protein disulfide isomerase (PDI) and tapasin (TPN) provide further stabilization. Peptides originated in the cytosol are translocated into the ER through the transporter-associated with antigen processing (TAP). The complex of TAP, TPN, MHC class I, ERp57 and CRT is called the peptide loading complex and ensures efficient peptide loading onto class I complexes. Once binding a peptide with sufficient affinity, the PLC disassembles and trimeric complexes are transported through the golgi to the cell

membrane. The contact MHC/peptide-TCR is the first step to build up a T cell immune response, which in the case of MHC-I culminates in death of the presenting cell by a specific cytotoxic CD8+ T cell.

Genes of the MHC, which are unique to the genome of jawed vertebrates, are key elements of cell-mediated immune responses, greatly impacting the way a host reacts to pathogens. In recent years, reports on bony fish MHC genes have been accumulating in the literature. Commercial relevance of some fish species used in aquaculture has been fuelling research in the field of fish immunology. European sea bass (*Dicentrarchus labrax*, L.) is a widely cultured species in Mediterranean fish farms. For this reason, a thorough characterization of its immune system is currently of extreme importance for understanding host-pathogen interactions and for vaccine development. These have been two major lines of research of the Fish Immunology and Vaccinology Group, where this thesis was developed. In this context, the working plan included cloning, sequencing and molecularly/structurally characterizing the presentation and processing genes of the class I pathway, for which there was no previous knowledge in this species.

The gene and cDNA of sea bass β2m (Dila-β2m) and several cDNAs of MHC-I HC (Dila-UA) were characterized. While Dila-β2m is single-copy, numerous Dila-UA transcripts were identified per individual with variability at the peptide-binding domain (PBD), and unexpected diversity at the cytoplasmic regions (CYT). Phylogenetic analysis places all Dila-UA sequences in the U lineage of teleost fish. The α1 domains resemble those of the recently proposed L1 trans-species lineage. Although no Dila-specific α1, α2 or α3 sub-lineages could be observed, two highly distinct sub-lineages were identified at the CP/TM/CYT regions. The three-dimensional homology model of sea bass MHC-I complex is consistent with other characterized vertebrate MHC-I structures. Furthermore, basal tissue-specific expression profiles were determined for both molecules, and expression of β2m was evaluated after poly I:C stimulus. Results suggest these molecules are orthologues of other β2m and teleost classical MHC-I.

TAP1 and TAP2 genes and transcripts were isolated and characterized. Only the TAP2 gene is structurally similar to its human ortholog. As other TAP molecules, sea bass TAP1 and TAP2 are formed by one N-terminal accessory domain, one core membrane-spanning domain and one canonical C-terminal nucleotide-binding domain. Homology modelling of the sea bass TAP dimer predicts that its quaternary structure is in accordance with that of other ABC transporters. Phylogenetic analysis segregates sea bass TAP1 and TAP2 into each subfamily cluster, placing them in the fish class.

The gene and cDNA of the glycoprotein TPN have been isolated and characterized. The mature form retains a conserved N-glycosylation site three conserved mammalian

tapasin motifs, two Ig superfamily domains, a transmembrane domain and an ER-retention di-lysine motif at the C-terminus, suggestive of a function similar to mammalian tapasins. A three-dimensional homology model of sea bass tapasin was calculated and is consistent with the structural features described for the human molecule.

CRT (Dila-CRT) gene and cDNA have been isolated and characterized. The mature protein retains two conserved motifs, three structural/functional domains (N, P and C), three type 1 and 2 motifs repeated in tandem, a conserved pair of cysteines and ER-retention motif. It is a single-copy gene composed of 9 exons. Dila-CRT three-dimensional homology models are consistent with the structural features described for mammalian molecules.

The last chapter reports the isolation and characterization of full cDNA and genomic clones from sea bass PDI (Dila-PDI) and ERp57 (Dila-ERp57). As described in other species, both molecules are composed of four thioredoxin-like domains (abb'a) followed by a C-terminal tail, retaining two CGHC active sites and an ER-signalling sequence, suggestive of a conserved function. Additionally, three-dimensional homology models further support Dila-PDI and Dila-ERp57 as orthologs of mammalian PDI and ERp57, respectively. Finally, high similarity is observed to their vertebrate counterparts (> 69% identity), especially among the few ones from closely related teleosts (> 79% identity).

In conclusion, transcripts and genes of MHC-I HC, β2m, TPN, TAP1, TAP2, CRT, ERp57 and PDI were identified following a homology cloning strategy. Collectively this work shows evolutionary conservation of the basic structures of these molecules, suggestive of similar functions. Furthermore, this work reports important primary data that can be used to develop new tools and further study the roles of these elements in sea bass immunity, particularly the class I presentation pathway and the establishment of CD8+ cytotoxic immunity.

Resumo

Todos os organismos necessitam de um sistema imune que os proteja de uma variedade de agentes invasores. Contudo, as ferramentas e os mecanismos disponíveis para levar a cabo uma resposta imune variam entre os diversos animais. Dependendo do nível de organização da espécie, vários mecanismos de defesa poderão ser activados em simultâneo. Os vertebrados, que surgiram há aproximadamente 500 Ma, incluem actualmente dois grupos: amandibulados (lampreias e peixes-bruxa) e mandibulados (peixes cartilagíneos e ósseos, anfíbios, répteis, aves e mamíferos). A particularidade destes grupos de organismos é o aparecimento de um novo tipo de célula imune - linfócito - que contém receptores de antigénio diversificados de forma somática e expressos de forma clonal, de tal maneira que cada linfócito expressa um receptor único. Como consequência, pode gerar-se um vasto repertório de receptores de antigénio/células imunes. Existem duas linhagens diferentes de linfócitos, B e T, com funções na resposta imune humoral (anticorpos) e celular, respectivamente. No caso dos vertebrados mandibulados, os receptores de antigénio têm por base domínios de imunoglobulina. Os receptores de células B (BCR) existem à superfície dos linfóctos B, enquanto os receptores de células T (TCR) estão presentes na superfície da membrana plasmática dos linfócitos T. As células T apenas reconhecem antigénios quando estes lhes são apresentados no contexto de moléculas do complexo maior de histocompatibilidade (MHC). Estas por sua vez estão divididas em MHC classe I (MHC-I) e MHC classe II (MHC-II), que respectivamente apresentam, de uma forma geral, péptidos endógenos a linfócitos T CD8+ e péptidos exógenos a linfócitos T CD4+. As moléculas de MHC-I existem em quase todas as células nucleadas, enquanto as de MHC-II estão presentes apenas nas chamadas células apresentadoras de antigénio (APC) profissionais. As células dendríticas que constituem o mais importante tipo de APCs, são na verdade uma inovação dos vertebrados mandibulados. Além da importante função na apresentação de péptidos, informando em que estado se encontra uma célula, as moléculas de MHC são também cruciais na educação dos linfócitos T, um processo que tem lugar no timo, um órgão linfóide também específico dos vertebrados.

As moléculas de MHC classe I são formadas por uma cadeia pesada (HC) e uma molécula de β2-microglobulina carregando um pequeno péptido. A 'montagem' é um processo estritamente regulado que ocorre no interior do retículo endoplasmático (ER) com intervenção de moléculas comuns e outras específicas. A flexibilização oxidativa inicial da HC por forma a atingir a conformação correcta envolve a calnexina e o ERp57. O dímero HC/β2m que subsequentemente se forma é altamente instável na ausência de

péptidos. As moléculas calreticulina (CRT), ERp57, PDI e tapasina (TPN) proporcionam-lhe estabilidade. Péptidos formados no citoplasma são transportados para o ER através do transportador associado ao processamento de antigénio (TAP). O complexo formado pelo TAP, TPN, MHC-I HC, β2m, ERp57 e CRT designa-se complexo de carregamento de péptido (PLC), e assegura o eficiente carregamento de péptidos aos complexos MHC-I. Uma vez ligado a um péptido com afinidade suficiente, o PLC é desagregado, e os complexos MHC-I triméricos são transportados através do aparelho de Golgi até à membrana celular. O reconhecimento MHC/péptido-TCR constitui o primeiro passo para que se desenvolva uma resposta imune com células T, a qual no caso do MHC-I culmina na morte da célula apresentadora pela célula T CD8+ citotóxica.

Os genes do MHC, que existem apenas no genoma dos vertebrados mandibulados, são elementos chave da imunidade celular, afectando largamente o resultado de uma doença. Recentemente, têm-se acumulado na literatura artigos referentes aos genes do MHC dos peixes ósseos. A relevância comercial de algumas espécies de peixes usadas em aquacultura tem incentivado a investigação na área da imunologia de peixes. O robalo (*Dicentrarchus labrax*, L.) é uma espécie altamente cultivada em aquaculturas do Mediterrâneo. Por esta razão, uma caracterização minuciosa do seu sistema imune é de extrema importância para que se compreendam interacções parasita-hospedeiro e para o desenvolvimento de vacinas. Estas são na verdade as principais linhas de investigação do grupo de Imunologia e Vacinação de Peixes do IBMC, onde o trabalho desta tese foi desenvolvido. Neste contexto, o plano de trabalhos incluiu a clonagem, sequenciação e caracterização molecular e estrutural dos genes de MHC-I envolvidos na apresentação e no processamento de antigénio, acerca dos quais nada se sabia nesta espécie.

O gene e o cDNA da β2m (Dila-β2m) e vários cDNAs da HC MHC-I (Dila-UA) do robalo foram caracterizados. Existe um único gene da Dila-β2m, mas vários transcritos de Dila-UA foram identificados por individuo, apresentando variabilidade no domínio de ligação ao péptido (PBD), e diversidade inesperada ao nível das regiões citoplasmáticas (CYT). A análise filogenética coloca as sequências Dila-UA na linhagem U dos peixes teleósteos. Os domínios α1 assemelham-se aos da linhagem trans-específica L1 recentemente proposta. Apesar de não terem sido identificadas sub-linhagens para os domínios α1, α2 e α3, foram encontradas duas sub-linhagens claramente divergentes nas regiões CYT. Os modelos tridimensionais de homologia do complexo MHC-I do robalo são consistentes com as estruturas de outros vertebrados. Os perfis de expressão basal foram determinados para ambas as moléculas em diferentes tecidos e a expressão da β2m foi também avaliada após um estímulo com poly I:C. Os resultados sugerem que estas moléculas são ortólogas de outras β2m e de MHC-I clássicos de teleósteos.

Os genes TAP1 e TAP2 foram isolados e caracterizados. Apenas o TAP2 tem uma organização idêntica ao equivalente humano. Tal como outras moléculas TAP, as moléculas TAP1 e TAP2 do robalo são formadas por um domínio N-terminal acessório, um domínio central que atravessa a membrana e um domínio C-terminal típico de ligação a nucleótidos. A modelação por homologia do TAP do robalo prevê que a sua estrutura quaternária é concordante com outros transportadores ABC. A análise filogenética separa o TAP1 e TAP2 do robalo em cada uma das subfamílias, colocando-os na classe dos peixes.

O gene e o cDNA da TPN foram isolados e caracterizados. Na sua forma matura, esta proteína retém um local conservado de N-glicosilação, três motivos típicos de TPNs de mamíferos, um domínio transmembranar e um motivo formado por duas lisinas responsável pela retenção no ER, sugerindo uma função semelhante às tapasinas de mamíferos. O modelo tridimensional da TPN do robalo calculado por homologia é consistente com as características estruturais descritas para a molécula humana.

O gene e o cDNA da CRT (Dila-CRT) foram isolados e caracterizados. A proteína contém dois motivos conservados, os três domínios estruturais/funcionais (N, P, C), três motivos tipo 1 e 2 repetidos em sequência, um par conservado de cisteínas e um motivo de retenção no ER. Os modelos tridimensionais obtidos por homologia são consistentes com as características estruturais descritas para as moléculas de mamíferos.

O último capítulo descreve a identificação e a caracterização de clones completos de cDNA e gene do PDI (Dila-PDI) e ERp57 (Dila-ERp57) do robalo. Tal como descrito noutras espécies, ambas as moléculas do robalo são compostas por quatro domínios de tioredoxina (abb'a') seguidos por uma cauda C-terminal, dois locais activos CGHC e uma sequência de sinalização do retículo, sugerindo uma função conservada. Os modelos tridimensionais determinados por homologia também suportam a ortologia das moléculas de robalo com as de mamífero. Finalmente, foi observada uma elevada semelhança com os equivalentes noutros vertebrados (> 69% identidade), especialmente os de outros peixes ósseos (> 79% identidade).

Concluindo, os transcritos e os genes das moléculas MHC-I HC, β2m, TPN, TAP1, TAP2, CRT, ERp57 e PDI foram obtidos seguindo um estratégia de clonagem por homologia. No conjunto, este trabalho mostra a conservação evolutiva das estruturas básicas destas moléculas, sugerindo funções similares. Além disso, os resultados descritos serão importantes para o desenvolvimento de novas ferramentas e futuros estudos que visem determinar o papel destes elementos na imunidade do robalo, em particular na via de apresentação de antigénio do MHC-I e no estabelecimento de respostas imunes mediadas por células T citotóxicas CD8+.

CHAPTER 1

General introduction

1. Comparative immunology

All organisms must be able to protect themselves from a variety of invading pathogens (including viruses, bacteria, fungi and parasites) in order to survive. Independently of their nature, pathogens will cause their hosts some level of fitness-diminishing harm, induce responsiveness, and undertake evasive strategies [1, 2]. The need for self-defence in the struggle for survival favoured the evolutionary development of simple-to-complex immune systems across the diversity of animal life (Fig 1). Depending on the organism, immune systems may comprise molecules, cells, tissues and organs, whose coordinated action (immune response) protects hosts from the infectious microbes (or even non-infectious foreign macromolecules) [1]. Individual species typically employ several mechanisms simultaneously to achieve functional redundancy of such an important system [3].

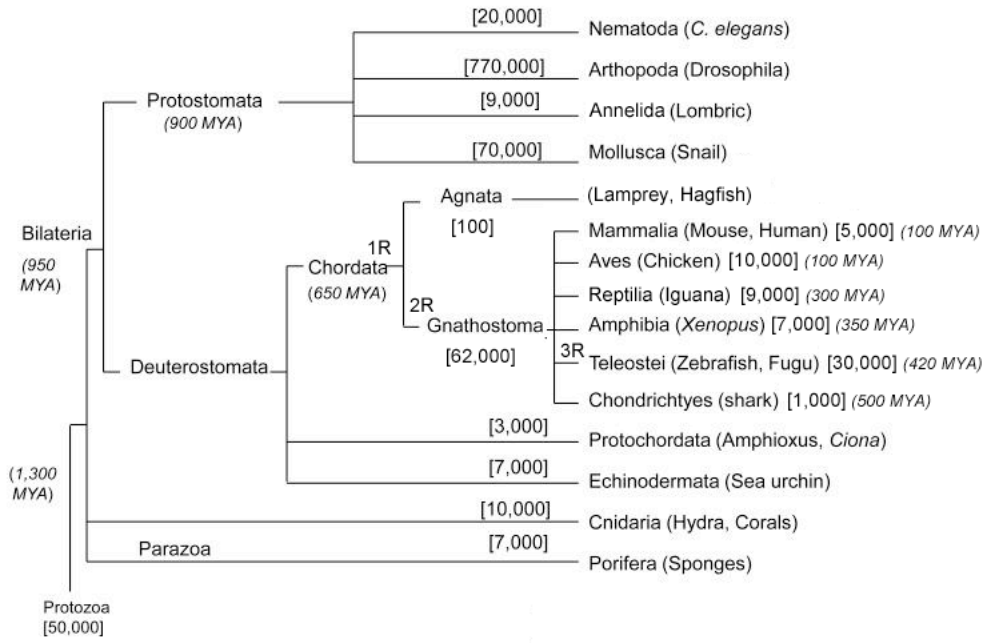


Figure 1. Simplified phylogeny of animals. Approximate time in millions of years for the emergence of the main group of organisms is in parenthesis, and approximate number of species of main taxa are indicated is in brackets. Examples of organism(s) of each class are given. Morphological and functional considerations and comparative genome analyses have led to the notion that whole genome duplications accompanied the evolution of vertebrates [4]. The first genome duplication (1R) is believed to have given rise to the ancestor of all vertebrates; the increased genome complexity provided the opportunity for morphological and functional diversifications. A second genome duplication (2R) is thought to have led to the emergence of jawed vertebrates. A third genome duplication (3R) has occurred in the stem lineage of ray-finned fishes, giving rise to a teleost-specific genome organization. Adapted from [5, 6].

1.1 Historical view

Historically, immune functions have been divided into two separate domains, with independent functions: innate and adaptive immunity. Innate defences, providing the first line of response, relied on programs of enzymatic (e.g. complement system) and cellular activation (e.g. macrophages) [3], underlying immune responses such as the recognition of pathogen-associated molecular patterns (PAMPs) by pattern recognition receptors (PRRs) [7] that rapidly induce antimicrobial responses. Adaptive responses, characterized by the capacity of memory formation, were attributed to the function of different types of lymphocytes [3]. This dichotomy was supported by their sequential appearance in evolutionary history, innate responses being already present in invertebrates, and adaptive responses being the hallmark of vertebrate immunity [3]. However, recent research led to the realization that innate immune systems also incorporate certain aspects of immune memory and that innate and adaptive immune functions are intimately linked [3]. Many of these insights emerged from comparative analyses in a limited number of species [3].

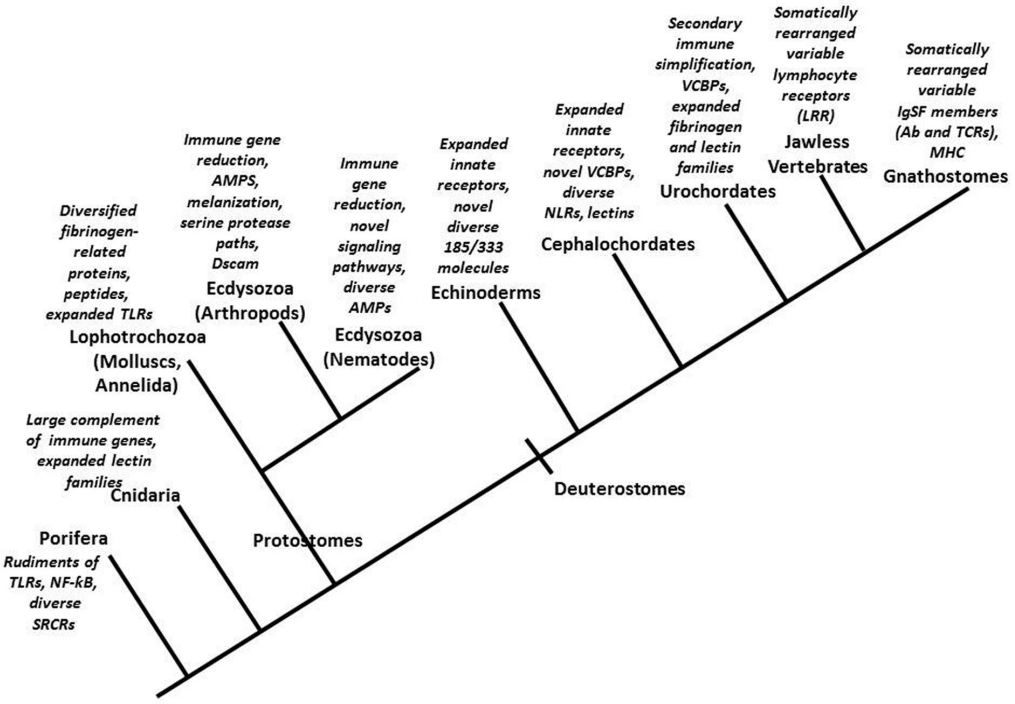


Figure 2. Overview of novel features of immune responses in distinct animal lineages. Recent findings are influencing the understanding of immune function and mechanisms of immune recognition. Four lineages are represented Porifera, Cnidaria, Protostomes and Deuterostomes. Relationships are not to time of divergence scale. TLR, Toll-like receptor; FREP, fibrinogen-related proteins; AMP, antimicrobial peptide; Dscam, Down syndrome cell adhesion protein; VCBPs, variable region-containing chitin-binding proteins; NLRs, intracellular NOD-like receptors; LRR, leucine-rich repeat; IgSF, immunoglobulin superfamily; Ab, antibodies; TCR, T cell receptor; MHC, major histocompatibility complex. RAG-1 and RAG-2 have been identified in sea urchin, but only a gene-like *RAG1* has been characterized in amphioxus. Adapted from [1].

Undoubtedly, humans and mice are intensively studied by immunologists and most knowledge about immune defences is coming from research in these models, but important discoveries made in non-model organisms greatly influenced the field of immunology [8]. E.g. observations of phagocytosis by amebocytes in echinoderms were in the origin of the concept of self/non-self recognition [9]; cellular and humoral immune functions in T and B lineages of lymphocytes [10] as well as gene conversion as a strategy in diversification of immune receptors [11] were firstly found in birds; studies in insects identified Toll receptors as mediators of innate immunity [12]; more recently, the ability of B cells to act as professional phagocytes came from studies in teleosts [13] and led to the discovery of phagocytic B cells in mammals. Clearly the study of other groups of organisms from invertebrates to vertebrate ancestors is relevant to our understanding of mammalian immunology.

1.2 The current picture

The emergence of new forms of immunity across animal life is currently a topic of great interest [1, 3, 14-16]. Invertebrates, comprising the majority of animal species on Earth (Fig. 1), can display simple-to-complex immune systems [1, 17] that are comparable to those of vertebrates [18, 19] in terms of the multitude of (i) mechanisms employed, (ii) immune-related molecules and (iii) effector cells (Fig. 2) [3]. Recent findings are detailed in Table 1. Discoveries go from observation of elements used in vertebrate immune recognition and microbial defence in sponges, to the realization that cnidarians have a complete set of immune genes in common with humans [1], but also from genome reductions in some organisms to expanded innate immune systems in others (Table 1) [1], giving the impression that a simplification of innate immunity occurred in mammals [2]. Additionally, as mentioned above, invertebrates also seem to exhibit hallmarks previously associated to vertebrate adaptive immunity [3]. Namely, several distinct mechanisms for generating diverse antigen receptors occur in molluscs (FREPs), arthropods (Dscam), echinoderms (Sp185/333) and cephalochordates (VCBPs) [1] (Fig. 2, Table 1). Furthermore, humoral and cellular responses seem also relevant in invertebrates [3]. Surprising is the observation that invertebrates can have increased and cross-protecting resistance to pathogens [20-22] with effects on immune effector cells [23]. This memory in invertebrates represents a state of alert [3] and has been referred to as ‘trained immunity’ [23], distinguishing it from vertebrate immunological memory based on clonal selection of immune cells [24]. In contrast to vertebrates, no antigen-triggered proliferative responses have yet been demonstrated [3]. Additionally, innate cell (NK cell) memory has been reported in mice [25].

Table 1. Overview of animal immune systems emphasizing distinctive features and diversified defence molecules (Adapted from [1])**.**

Phylum	Body plan	Specialized defense cells	Genome reduction	Distinctive immune features	Diversified defense molecules
Porifera (sponges)	Diploblastic	No	No	TIR domains, lack external LRR, MyD88 homolog, NF-κB rudiments, lack death domains, LPS-interacting proteins, perforin-like molecules, antiviral 2′-5′ oligoadenylate system	Scavenger receptor cysteine-rich molecules
Cnidaria (anemones, corals, jellyfish, *Hydra*	Diploblastic with mesoglea	No	No	Bonafide TLR and NF-κB pathway, complement-3 component, multiple NACHT domains and NLRs likely, recognizable RAG1 homolog. *Hydra* lacks canonical TLR with both LRR and TIR, allorecognition molecules present including with IgSF domains	Diverse C-type lectins
BILATERIA PROTOSTOMES ECDYSOZOA					
Nematoda (*Caenorhabditis* and many others)	Triploblastic	No	Yes	One TLR that plays a role in defense against some bacteria, lacks canonical Toll pathway and NLRs, but can mount inducible defense responses and have several novel defense-related signaling pathways	Produces many caenopores and other antimicrobial peptides including 42 NLPs caenacins, diverse C-type lectins
Arthropoda (insects, crustaceans, and many others	Triploblastic	Yes	Yes	NLR and complement-3 components lacking, have one or two TLRs functioning with NF-κB pathways, TLRs do not engage ligands directly, produce antimicrobial peptides, CLIP-protease cascades, melanization reactions	Multimeric fibrinogen-related molecules, the IgSF member Dscam with multiple isoforms
LOPHOTROCHOZOA					
Annelida (earthworms, leeches, polychaetes)	Triploblastic	Yes	No	Over 100 TLR genes, extensive involvement of coelomocytes from coelom in defense, cytotoxicity against allogeneic cells, hemolytic and clotting factors in body fluid, antimicrobial peptides, and protective body mucus	Expanded set of TLRs in polychaetes
Mollusca (cephalopods, snails, bivalves chitons, others)	Triploblastic	Yes	No	Involvement of body mucus in protection, TLRs and Toll pathway present, little melanization, hemocytes working with lectins like galectins or fibrinogen-containing proteins (FREPs), mitogen-activated protein kinase pathways, complement-like factors antimicrobial peptides	Somatic diversification of FREPS by point mutation and gene conversion, large C-type lectin families, diversified myticin C in bivalves
DEUTEROSTOMES					
Echinodermata (sea urchins, starfish, brittle stars, crinoids, sea cucumbers)	Triploblastic	Yes	No	"Expanded" innate immune system with >220 TLRs, >200 NLRs, >200 SRCR genes, Toll pathway, lectin and alternative pathways, RAG1 and RAG2 homologs present	In addition to expanded sets of TLRs, NLRs, and SRCRs, they also have novel Sp 185/333 gene family producing diverse immune proteins
CHORDATA					
Cephalochordata (amphioxus)	Triploblastic	Yes	No	"Expanded innate immune system with ∼72 TLRs, >92 NLRs, −270 SRCRs, >1200 C-type lectins, possesses distinctive variable region-containing chitin-binding proteins (VCBPs), have functioning complement, RAG1 and possibly RAG2 present	In addition to expanded sets of TLRs, NLRs, SRCRs, and lectins, they have polymorphic VCBPs arising from a variety of mechanisms
Urochordata (tunicates)	Triploblastic	Yes	Yes	V-like and CI-like domains present, VCBPs present, have three TLRs, lack complement, or expansion of any gene family relevant to vertebrate immunity	Have expanded families of C-type lectins and fibrinogen-related proteins
Agnathans (lampreys, hagfish)	Triploblastic	Yes	No	Lack RAG1 and RAG2 and do not produce TCRs or immunoglobulins, but do have two basic types of lymphocytes and produce variable lymphocyte receptors (VLRs) with LRRs	Produce diverse VLRs through somatic rearrangement of modules with leucine-rich repeats
Gnathostomes (fish, amphibians reptiles, birds, mammals)	Triploblastic	Yes	No	Modest numbers of TLRs (10–25) and NLRs (20–35), three complement pathways, somatic diversification of both Ig and TCR, involvement of MHC, memory, heightened secondary response, affinity maturation in some	Both Ig and TCRs diversified somatically

1.3 Important observations

According to Litman and Copper [8], some of the most notable discoveries recognized on the evolution of immunity include the following: (i) innate immunity preceded adaptive immunity in the evolution of immune recognition, providing the common basis of immune recognition between invertebrates and vertebrates [26, 27]; (ii) invertebrate immune-type receptors may undergo diversification [28, 29], and even immunological memory and cross-resistance can be achieved [22]; (iii) polymorphic cellular receptors in marine invertebrates may ensure self/non-self recognition of cells [30]; (iv) clonally diverse lymphocytes are a vertebrate innovation [8].

2. Vertebrate immune system

Vertebrates are divided in two groups: jawless (Agnatha) and jawed (Gnathostomata) vertebrates, comprising roughly 100 species of lampreys and hagfish and approximately 62,000 species from cartilaginous fish to mammals [6], respectively. It has been suggested that the emergence of vertebrates involved a change in the structure/function of the ancestral immune system [14, 15, 18, 19, 31, 32]. Normally, it is highlighted that vertebrates added adaptive immunity to innate defences, gaining antigen-specific memory and more rapid and efficient secondary immune responses [32]. However, as a consequence of the recent findings mentioned above, the functional distinctions between innate and adaptive immunity are becoming less clear [3]. Despite these developments, the vertebrate immune system exhibits unique innovations [3], further discussed below.

2.1 Two different systems

Lymphocyte-based adaptive immunity emerged twice in the early phases of vertebrate evolution (convergent evolution) demonstrating clear selective advantages of being equipped with such an immune system [32]. Uniquely, both jawless and jawed vertebrates have an adaptive immune system based on somatically diversified and clonally expressed antigen receptors (Fig. 3) [18]. With somatic diversification (Box 1), a wide range of antigen specificities is created, requiring elimination of unwanted self-reactivity [33]. The fact that each lymphocyte is armed with a single receptor [18] confers an advantage regarding the elimination of the repertoire of self-reactive cells and generation of immune memory [3]. Despite having common design principles in generating a vast repertoire of antigen receptors/cells, the molecular nature of the antigen receptors and their mechanisms of assembly are different in the two clades [32]. Furthermore, besides the emergence of novel cell types (lymphocytes and also dendritic cells) and distinct genetic innovations (VLRs vs Ig/TCR/MHC), novel immune-related tissues (thymus and spleen) are also vertebrate-specific [32]. Functional cooperation of innate and adaptive arms is fundamental [32], and many of the effectors operating this crosstalk already exist in ancient vertebrates [32].

Remarkably, all vertebrates possess two major types of lymphocytes, B-like and T-like cells, mediating humoral and cellular immunity, respectively (Fig. 3) [34]. B cells express antigen receptors on their surface and upon antigen encounter can secrete them as antibodies. In contrast, the antigen receptors of vertebrate T cells are always cell surface-bound, even after antigenic stimulation [32].

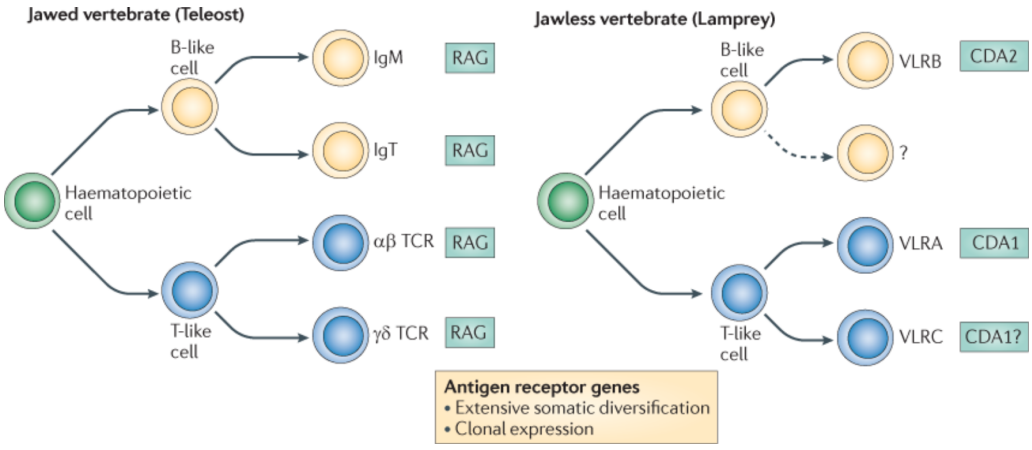
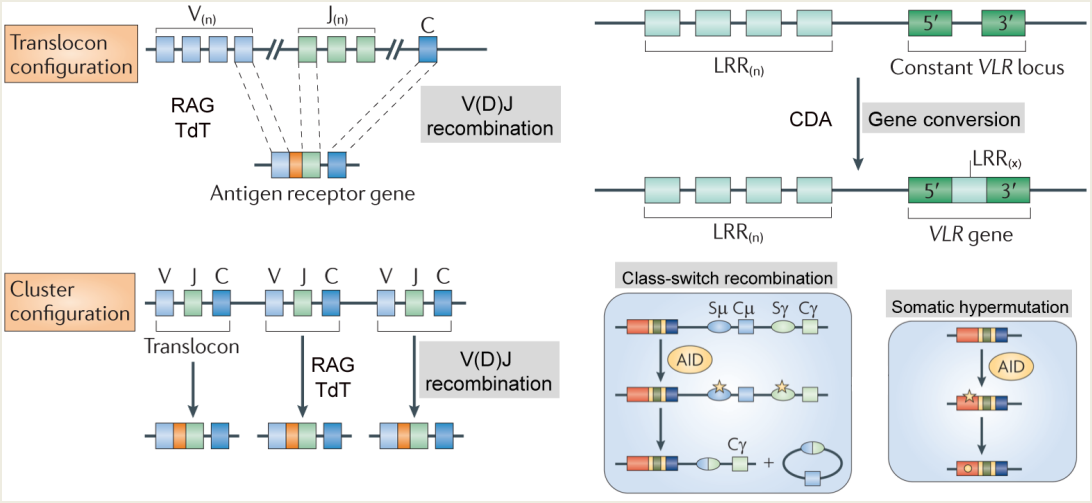


Figure 3. Specificities of lymphocytes. Both groups of vertebrates possess T- and B-like cells. In jawed vertebrates, B and T cells have two sub-lineages each. This seems also true for the T-like cell lineage of jawless vertebrates, but is currently unknown if a second B-like cell lineage exists in jawless fish. In jawed vertebrates, B cells express immunoglobulins (Igs) and T cells express T cell receptors (TCRs). Distinct sub-lineages of lymphocytes express structurally different antigen receptors. However, for B cells this is not always the case; for instance, all mammalian B cells express the same type of Ig (not shown). In lampreys, there are three different types of variable lymphocyte receptors (VLRs): VLRB is expressed by B-like cells, while VLRA and VLRC are expressed by separate lineages of T-like cells. All these antigen receptors are subject to somatic diversification. For VLRs, this occurs by gene conversion, which is presumably initiated by cytidine deaminases (CDAs), and the two known CDA genes (*CDA1* and *CDA2*) are expressed in a lineage-specific manner. In jawed vertebrates, somatic diversification of Igs and TCRs occurs by V(D)J recombination, which is carried out by the recombination activating gene (RAG) proteins and associated molecules. All vertebrate antigen receptors are clonally expressed. Adapted from [3].

2.1.1 Jawless vertebrates

Jawless vertebrates rely on leucine-rich repeat (LRR)-containing antigen receptors termed variable lymphocyte receptors (VLRs) [19, 35]. Three VLR genes have been identified: VLRA, VLRB and VLRC [18]. VLRB antigen receptors are expressed by the B-like cells of lampreys and can be secreted, similar to antibodies [36]. By contrast, VLRA and VLRC antigen receptors [35, 37] are exclusively expressed by two distinct T-like cell lineages [37, 38] and are located on the cell membrane, similar to a T-cell receptor (TCR) [39]. VLRs are assembled from incomplete genomic elements by a gene conversion process [40] (Box 1), with the involvement of cytidine deaminases (CDA) [41]. The two known CDAs are expressed in a lineage-specific fashion: CDA1 is involved in VLRA (and possibly VLRC) assembly, while CDA2 participates in VLRB assembly [35]. The diversity of VLRs is generated by combinatorial use of variable numbers of leucine-rich repeat (LRR) elements [19, 42].

Box 1. Somatic diversification



Adapted from [14, 18].

Changes in the sequence of DNA that occur in individual cells and their progeny. Includes V(D)J recombination, gene conversion, somatic hypermutation and class switch recombination.

V(D)J recombination

Generates highly diverse repertoires of T and B cell receptors (TCR and BCR). Variability is achieved by combination of pre-existing multi-copy gene segments called variable (V), diversity (D) and joining (J) segments found in the BCR-heavy and TCR-β and -δ chains. D segments are absent in the BCR-light (κ and λ) and TCR-α and -λ chains, so only V and J recombination occurs. It involves breakage and repair of DNA [14]. The reaction is initiated by the recombination activation gene 1 (RAG1) and RAG2 proteins, that bind to conserved recombination signal sequences (RSS) flanking all recombining gene segments in Ig and TCR loci [33]. The recombinase introduces double-stranded DNA breaks [14]. Subsequently, terminal deoxynucleotidyltransferase (TdT) adds random nucleotides to gene elements (junctional diversity), dramatically increasing repertoire diversity [14]. Finally, fusion of segments, i.e. non-homologous end-joining, takes place with the help of DNA repair proteins. The RSSs are joined without further end processing and form excision circles [14]. This recombination process involves a loss of germline-encoded nucleotides and addition of non-templated nucleotides [33]. It occurs in all jawed vertebrates [33] either in translocon or cluster configuration. Combinatorial diversity and junctional variability synergize to generate receptor specificities [33]. Once functional DNA rearrangements occur, TCR sequences are unaltered; however, after encounter with antigen, B cells further recombine the receptor by somatic hypermutation and class-switch recombination [14].

Gene conversion

This is a process during which heteroduplexes between non-identical stretches of DNA are formed and resolved through recombination [33]. It is initiated by DNA double-strand breaks and results in nonreciprocal transfer of genetic information [19], i.e. the donor gene(s) remains unmodified and an acceptor gene acquires the recombined segment. In chickens, variable (V)-region pseudogenes are donors that modify the functional, rearranged V gene [43], and this process generates a diverse repertoire. The generation of functional variable lymphocyte receptor (VLR) genes in lampreys also uses a similar process. In VLR genes, multiple short stretches of sequences encoding leucine-rich repeats (LRRs) are situated next to the single incomplete constant part of the VLR locus [19]. Gene conversion generates one

complete gene per allele, similarly to the translocon configuration [18]. This mechanism does not generate junctional diversity [33].

Somatic hypermutation

Process that starts upon receptor-antigen interaction and targets rearranged V(D)J regions of immunoglobulin genes [33]. Its purpose is to provide a clonal population of B cells expressing receptors with higher affinity for an antigen. It is initiated by activation-induced cytidine deaminase (AID), which deaminates individual cytidines within the V(D)J exon of the Ig gene, leading to uracyl (U):guanidine (G) mismatches [14]. Subsequent error-prone repair results in individual point mutations and B cells with higher affinity for the original antigen are selected [14].

Class-switch recombination

This process exchanges the constant region domains of a rearranged Ig gene by a deletional DNA rearrangement [33]. Since each constant region mediates a specialized function, switching permits adaptive guidance of antibodies, providing alternative effector functions [33]. Creating a new heavy chain requires loop-out and deletion of DNA between switch regions, employing transcription of the switch regions. Requisite switching factors include AID and components of general DNA repair [33]. AID creates U:G mismatches in the highly repetitive switch (S) regions that are upstream of the exons encoding the constant regions of different isotypes [14]. Error-prone repair leads to the generation of double-strand DNA breaks, excision of the intervening DNA and fusion of the remains of the switch regions [14].

2.1.2 Jawed vertebrates

The antigen receptors of jawed vertebrates belong to the immunoglobulin (Ig) superfamily. Igs serve as the receptors for B cells (BCR), whereas T cells use T-cell receptors (TCRs) [6]. Two functionally distinct lineages of T cells exist in jawed vertebrates, one expressing an αβ T cell receptor (TCR) and the other expressing a γδ TCR [44]. BCRs are composed of two membrane-spanning heavy chains and two associated light chains in the case of Igs, and TCRs of two membrane-spanning molecules, either α and β or γ and δ [32]. T-cell antigen receptors are further associated with cell-type specific CD8 or CD4 co-receptors. Functional antigen receptor genes are formed through V(D)J recombination (Box 1) initiated by recombination-activating gene (RAG) products; variable (V)-, diversity (D)-, and joining (J)-type segments from germline Ig and TCR genes are assembled to form functional antigen receptor genes [32]. In addition, non-germline-encoded sequence variability at the junctions (typically encoding the antigen-binding surface of the antigen receptors) makes a significant contribution to the diversity of functional BCRs and TCRs [32]. Junction diversity is considered an important difference between jawless and jawed vertebrates [3]. Cytidine deaminases are also used during generation of the primary repertoire of Igs in jawed vertebrates [43] and in their diversification during immune responses [45].

2.2 Myeloid cells as antigen presenting cells

Hematopoietic cells originate both the myeloid and lymphoid cells. Immune systems initially relied on multi-function innate cells from the myeloid lineage, with phagocytosis of pathogens and infected/damaged cells as a major function [32]. Lymphocytes represent an innovation of vertebrates [3] and include 'innate' (natural killer cells) and 'adaptive' lymphocytes (T and B cells), with sequential evolutionary appearance. T and B lymphocytes, but not NK cells, express somatically assembled antigen receptors, in a 'one receptor/one cell' fashion [3]. In vertebrates, the myelomonocytic cells (macrophages and dendritic cells) also function as antigen-presenting cells [32]. Importantly, they provide information to B and T lymphocytes either indirectly, through the production of chemokines and other cytokines [46], or directly presenting antigens through cell-to-cell contacts (Fig. 4) [47]. DCs are present in early jawed vertebrates [32]: they have been identified in teleosts [48-51] and their presence suggested in cartilaginous fish [52, 53]). Although not yet described, it is expected that jawless vertebrates also possess antigen-presenting cells and molecules [18]. The antigen presentation system of jawed vertebrates is discussed in section 3.

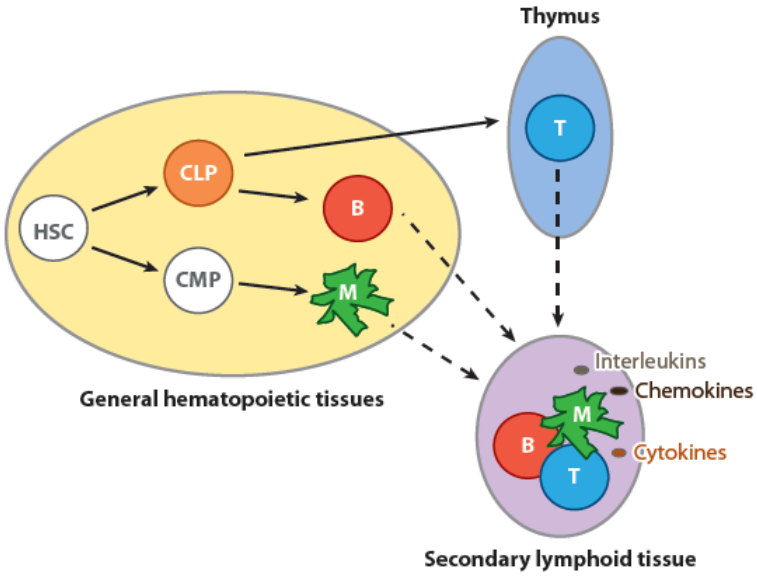


Figure 4. Developmental origin and functional cooperation of major immunological cell types. The precursors of lymphoid and myeloid lineages are believed to differentiate from hematopoietic stem cells (HSCs) in general hematopoietic tissues. The common lymphoid progenitor (CLP) gives rise to T and B lineages. Whereas T cell precursors migrate to and differentiate in the thymus (thymoid in lamprey), B cells mature in hematopoietic tissues. The common myeloid precursor (CMP) originates different cell types, including antigen-presenting dendritic cells (here indicated by M). The reciprocal interactions of lymphocytes and antigen-presenting cells through direct cell-cell contacts and/or via interleukins, chemokines and other cytokines regulate immune responses in secondary lymphoid tissues. Adapted from [32].

2.3 Lymphoid organs

Lymphoid organs are integral parts of vertebrate adaptive immune systems for ~500 million years (Fig. 5) [6]. Functionally they are divided into two main categories: (i) primary

or central and (ii) secondary or peripheral lymphoid organs [54]. The first are required for the development of lymphocytes and of their primary repertoire [54]. The second are involved in the coordination of immune responses, enabling interactions between immune cells [54]. Lymphoid tissues functionally equivalent to secondary lymphoid organs often emerge under inflammatory conditions and are in this case named tertiary lymphoid tissues [54].

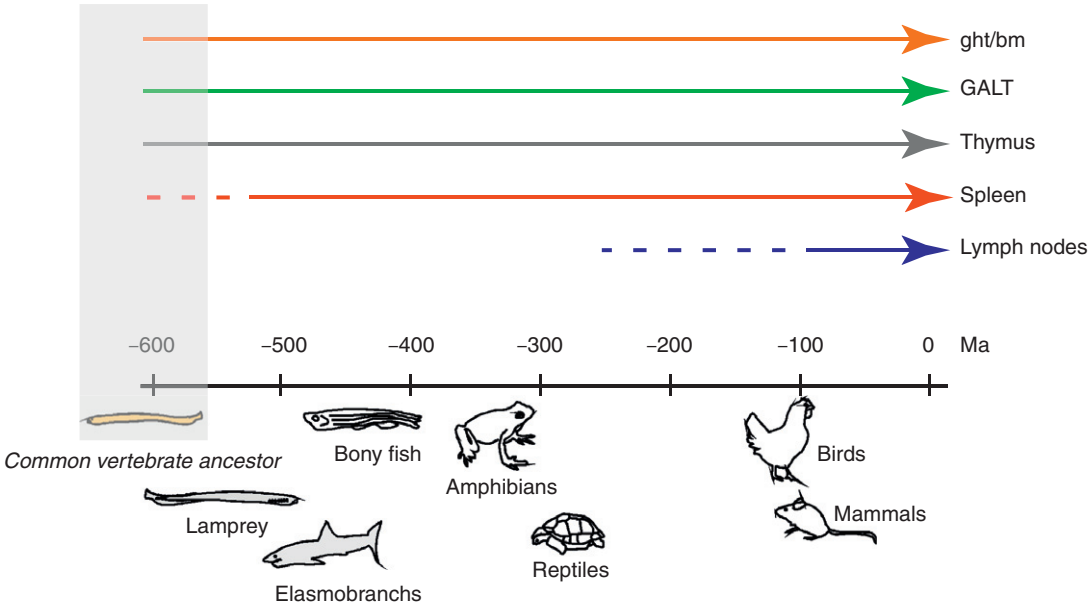


Figure 5. Evolution of lymphoid tissues in vertebrates. The emergence of vertebrate classes is indicated on a time scale of millions of years (Ma). General hematopoietic tissues (ght) in lower vertebrates are considered homologous to the bone marrow (bm) in tetrapods. The spleen probably originated in the common vertebrate ancestor. The evolutionary emergence of lymph node-like structures began with reptiles (dotted line). The phenotype of the extinct common ancestor of all vertebrates is hypothetical. Adapted from [6].

The sites of B lymphopoiesis vary among species according to the anatomical locations of general hematopoietic tissues, and successive stages of B-cell development take place in various locations (either within a specific tissue or in different tissues) [6]. For instance, in birds, B-cell development starts in the spleen [55], ending in the bursa of Fabricius [10], whereas in mammals the spleen appears to be the site where B cells complete the maturation process [56] that begins in the bone marrow.

T cell progenitors [57, 58] also originate in general hematopoietic tissues, but migrate to the thymus [59] for differentiation. Hence, in stark contrast to B cells, T-cell development is initiated and terminated in the thymus, which is the most ancient primary lymphoid organ and is present in all vertebrates (including lampreys, in the form of thymoids [60]) [6]. The thymus is always associated with the pharynx, although its precise location and organization varies considerably among vertebrate species [6].

Box 2. Fundamental features of bony fish immune systems.

	Teleosts	Mammals
Immunoglobulin	IgM, IgD and IgT (or IgZ)	IgM, IgG, IgA, IgD and IgE
AID	Yes	Yes
Class-switch recombination	No	Yes
Somatic hypermutation	+++	+++
Affinity maturation	+	+++
Memory responses	+	+++
TCR, CD4, CD8	Yes	Yes
MHC class I and II	Yes	Yes
CD28, CD40, CD80, CD86, ICOS	Yes	Yes
T_H1, T_H2 and T_H17 cytokines	Yes	Yes
Spleen, thymus and bone marrow	Spleen and thymus but no true bone marrow	Yes
Mucosa-associated lymphoid tissue	Yes	Yes
Germinal centers and lymph nodes	No	Yes

Comparison of key elements of immunoglobulin-based adaptive immune systems of teleost fish and mammals. MHC, major histocompatibility complex; ICOS, inducible costimulator; T_H1, T_H2 and T_H17, subsets of helper T cells.

The immune system of bony or teleost fish includes most of the elements of the innate immune system present in mammals: cytokines, lectins, complement system, PRRs (such as Toll-like receptors, TLRs; NOD-like receptors, NLRs; RIG-like receptors, RLRs; C-type-lectin receptor, CLRs; antinmicrobial peptides (AMPs), macrophages, neutrophils, NK cells [61, 62]. Several of these innate immune components, such as lectins, complement and natural killer cell receptors have a more diverse repertoire in teleosts than in mammals [63-65]. Importantly, they are among the earliest organisms with a lymphocyte-based adaptive immune system relying on rearranging immunoglobulin (Ig) segments through RAG gene products. Hence, they have Igs (B cell receptors/antibodies), T cell antigen receptors (TCR), MHC class I and II molecules, T and B cells, spleen and thymus [31]. Besides the similarities between teleost and mammalian immune systems, but specificities have also been reported. For instance, they have dendritic cells, the most specialized antigen-presenting cells, but lack lymph nodes and bone marrow, which are very important lymphoid tissues [66]. Nevertheless, they do have a functional ortholog of mammalian bone marrow, the head kidney, which represents the main teleost hematopoietic lymphoid tissue [66]. Teleosts also lack germinal centers and antibody class-switch recombination, despite expressing the activation-induced cytidine deaminase (AID). The mucosa-associated lymphoid tissue, that plays a crucial role in the maintenance of mucosal homeostasis, is found within the skin, gill and gut [67]. Only three immunoglobulin classes have been identified in teleosts: immunoglobulin M (IgM), IgD and IgT (or IgZ in some species) [68-70]. Representatives of the four mammalian TCR chains (α, β, γ and δ) have been described in teleosts [71].

Adapted from [72].

Notably, vertebrate evolution is characterized by an increasing complexity of secondary lymphoid tissues, suggested from the growing complexity of the spleen and from the emergence of novel structures, such as lymph nodes in more recent groups of organisms [32].

In bony fish, immune-related tissues include the mucous skin, gills, gut-associated lymphoid tissue, the mentioned thymus and spleen, and also the head-kidney [73]. Other important features of the teleost immune system are summarized in Box 2.

3. The jawed vertebrate antigen presentation system

3.1 Functions

In jawed vertebrates (cartilaginous and bony fish, amphibians, reptiles, birds and mammals), antigen presentation occurs through two pathways, ending with the formation of MHC-peptide complexes at the cell surface [74]. Peptides derived from intracellular degradation processes are loaded and the complexes delivered to the cell surface [75]. Providing processed antigen structures to T lymphocytes is the first step to initiate antigen-specific immune responses, since TCRs only recognize antigens in complex with MHC molecules [33]. On the other hand, the wide range of TCR antigen specificities created with somatic diversification requires a dedicated quality control system to eliminate self-reactivity [33]. Hence, as pivotal elements of this quality control system, MHC complexes are necessary for T-cell development in the thymus and for immune surveillance in the periphery [57]. Developing T cells are first subjected to positive selection to ensure that the TCR can interact with self-MHC molecules and in a subsequent negative-selection step, TCRs with high affinity for self are eliminated [33]. It has been hypothesized that such a quality control system was not created *de novo* but instead co-opted an ancestral recognition system discriminating between self and non-self, with expected function in sexual selection [33, 76]. Indeed, MHC-associated behavior in animals indicating that MHC peptide ligands act as olfactory cues, has been described in different species attesting involvement of MHC in sexual selection functions [76-80].

Of note, the chromosomal MHC region is unique to the genomes of jawed vertebrates. Hence, if a functionally equivalent antigen presentation system exists in lamprey and hagfish, as it is anticipated to, its molecular components must be different from the MHC system [32].

3.2 History and genetics

MHC research started in the beginning of the 20th century, when Little and Tyzzer [81] analysed the fate of tumours transplanted between mice. However, it was only in the late 1930's that Gorer identified the genetic factors influencing transplant outcome in mice [82]. George Snell further characterized these genetic factors in transplantation studies using congenic mouse strains (in two congenic strains, the mice are identical at all loci except the one at which they are selected to be different). In 1948 he proposed to call the genes responsible for causing a grafted tissue to be perceived as similar or different from one's own tissues, 'histocompatibility (H) genes', as they determine tissue compatibility between individuals [83]. The genetic region controlling graft rejection and containing

several linked genes was called major histocompatibility complex, or MHC.

Due to whole genome duplication events, the MHC occurs in four paralogous regions in the typical vertebrate genome, yet antigen presentation genes occur in only one paralogous segment [84]. Characteristically, the genes responsible for antigen presentation are: (i) polygenic - exist as multiple loci; (ii) polymorphic - multiple variable alleles are present within a population; and (iii) co-dominantly expressed - for each locus, alleles inherited from both progenitors are expressed.

3.3 Genomic organization

In humans, MHC or HLA (human leukocyte antigen) genes are closely linked in a region of approximately 4 megabases (without extended regions), located on chromosome 6 [85]. This is a gene-rich region including over 260 genes, 40% of which are associated with immune functions [86]. The MHC region is further divided into regions with similar functions, and includes (in order of physical location from centromere to telomere): extended class II, class II, class III, class I, and extended class I regions (Fig. 6) [87]. The number of genes and the presence of each region varies between species [87].

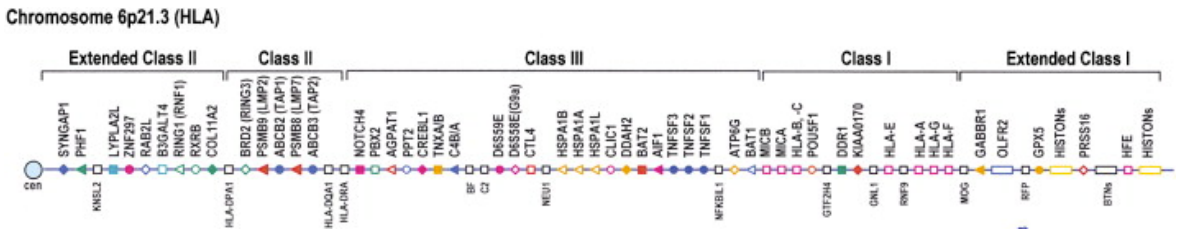


Figure 6. Human MHC region. Genes in human MHC are arranged from the centromere (cen) to telomere (map not drawn to scale). Human antigen processing genes such as immunoproteasome (LMP) and transporter (TAP) and LMP genes are located in class II region and TPN in extended class II region. The size of the extended MHC depicted here is approximately 8 megabases. From [84].

The class I region is composed of classical (Ia) and non-classical (Ib) genes. Classical class I molecules are expressed on nearly all nucleated cells and form heterotrimers consisting of one α or heavy chain (HC) and β2-microglobulin (β2m) coupled with a peptide. Of note, the β2m gene is located outside the MHC region in all jawed vertebrates except sharks [88], probably indicative of primitive sinteny. Typically, class Ia molecules present peptide antigens to CD8+ T lymphocytes, while the functions and expression profiles of the non-classical genes are diverse [87]. Members of both categories may also be monitored by natural killer cells, acting as ligands for their innate receptors [89, 90]. Class II molecules have their expression restricted to fewer cell types

that belong to the myeloid/lymphoid lineages, namely macrophages/monocytes, dendritic cells and B cells, and also form heterotrimers of one α chain, one β chain and the loaded peptide. These complexes present peptide antigens to CD4+ T lymphocytes. The class III region contains a highly dense selection of diverse immune (TNF family members, complement, etc.) and non-immune genes [87].

In teleosts, the genomic organization of MHC genes is different and variable patterns are found at the species level [84]. Uniquely, their class Ia and II genes are found in different linkage groups on different chromosomes [84]. For this reason, Stet proposed to call them simply MH genes instead [91]. In addition, as opposed to human, the antigen processing genes LMP (proteasome), TAP (transporter) and TPN (tapasin) are closely linked to class Ia genes, comprising a true class I region [84]. This holds also true for all non-mammalian vertebrates studied to date [84], such as chicken [92], amphibians [93] and cartilaginous fish [94]. The fact that Chondrichthyes possess linkage of some MHC class I, II and III genes [94], suggests that the common ancestor of all jawed vertebrates should possess an organized MHC. Finally, some non-immune genes found in the human class III region and in the extended class II region are also linked to class I in teleosts; but none of the class III immune genes is linked to the class I or class II [84].

3.4 Diversity of MHC genes

The number of class Ia loci varies between organisms. Most mammals possess and express two or more class Ia loci. In humans, the class I heavy chains are encoded by three genes (*HLA-A*, *HLA-B* and *HLA-C* in humans) [74]. HLA-A and HLA-B display higher expression levels than HLA-C [95]. The chicken MHC (or *B* locus), having only 19 genes dispersed over a 92 kb region after loss of non-essential elements, is characterized as the minimal essential MHC [92], still object of recombination [96]. Chicken have only two class Ia genes but only one is highly expressed at the cell surface [84]. There is also a single polymorphic class II β chain expressed while class II α is monomorphic [84]. Similar to chicken, *Xenopus* has one highly expressed class Ia molecule and two or three class II α and β [84]. Within teleosts, given the high number of species and the divergence time separating them, the scenario is quite variable. For instance, salmonids such as rainbow trout and Atlantic salmon, display also a single expressed class I α chain [97-99], but Atlantic cod has many (up to 43 loci) class I molecules [100]. Surprisingly, it has been found that Atlantic cod has lost the complete set of MHC class II genes [101]. Intermediate situations have also been described including e.g. two class Ia genes in medaka [102] or twelve in stickleback [103].

In contrast to human, chicken TAP and TPN genes have high allelic polymorphism

[104], while in *Xenopus* at least two allelic lineages of TAP have been described [105].

All teleost MHC class I (Ia and Ib) loci described to date have been classified into four different lineages: U, Z, L, and S [106-108]. The U lineage is present in all teleosts and includes both class Ia (highly polymorphic) and Ib (non-polymorphic) loci. The single expressed highly polymorphic locus of salmonids, as well as most of the teleost class I published sequences belong to the U lineage. The Z lineage is present in fewer but still numerous teleosts and includes genes with features typical of Ia genes, (such as ubiquitous expression and almost complete conservation of anchor residues) [107]. The L lineage, found only in cyprinids and salmonids, includes non-classical but extensively diversified and divergent genes [107]. Finally, the S lineage was described only in salmonids and catfish [107].

Notably, differences in genomic patterns observed between teleost species, include: conservative number of class I genes (limited diversity) in some teleosts, but higher number in others; conservative number of class II loci, with reported absence in one species; presence of different lineages of class I genes; different relative evolution rates between class I and class II regions (fragmented MHC allows class I and class II genes to evolve independently, under different selection pressures).

The different numbers of genes and varying organizations of MHC gene clusters can be explained by rapid evolution [87]. After divergence from the common ancestor, the MHC changes rapidly through repeated gene duplication (birth and death model [109, 110]) and polymorphism (point mutations - substitutions, insertions, deletions; and translocations - exon shuffling, gene conversion, unequal cross-over [111]), showing variation even among closely related species of the same lineage [87, 112]. On the other hand, standing genetic variation allows faster evolutionary changes (e.g. allele frequency shifts), as was recently demonstrated in two stickleback groups under the effect of distinct parasites [113]. Different pathogen pressures (pathogen-driven selection) [114], specific to each species, could drive rapid diversification through selection of resistant alleles [113] either by overdominance (heterozygous advantage) or negative frequency (rare allele advantage) [114]. MHC-based mate choice (sexual selection) also helps explain standing variation, although it is still unclear whether individuals aim at reducing inbreeding, select for specific MHC alleles, ensure offspring are heterozygous at locus or maximize diversity across several loci [114]. Another model explains variation not directly under selection, proposing that recessive disease-causing mutations can accumulate in the MHC and natural selection fails to purge them because they are only rarely expressed in a homozygous state (given the high number of alleles and gene diversity) [115]. Controversies and difficulties in evaluating their interacting roles apart, these mechanisms

are not mutually exclusive and can all contribute to the maintenance of high levels of MHC polymorphism [116].

3.5. Molecular structure of MHC

As previously mentioned, classical MHC-I molecules are involved in the formation of trimeric complexes consisting of a HC non-covalently associated to soluble β2m and a loaded peptide [117]. The HC consists of three extracellular domains (α1, α2, and α3), a membrane-anchored region and a cytoplasmic tail. The polymorphic residues of class I molecules are confined to α1 and α2 domains, which form a peptide-binding groove called the peptide-binding region (PBR). This consists of two anti-parallel α-helices forming the walls and a platform of eight anti-parallel β-strands (Fig. 7). The ends of the class I peptide-binding groove are closed, limiting the maximum length of class I-binding peptides to 8-11 amino acids. The membrane-proximal α3 domain and β2m are both homologues with the Ig C1 domains. CD8 co-receptor (on αβ T cells) interacts with a conserved region of the α3 domain of HC.

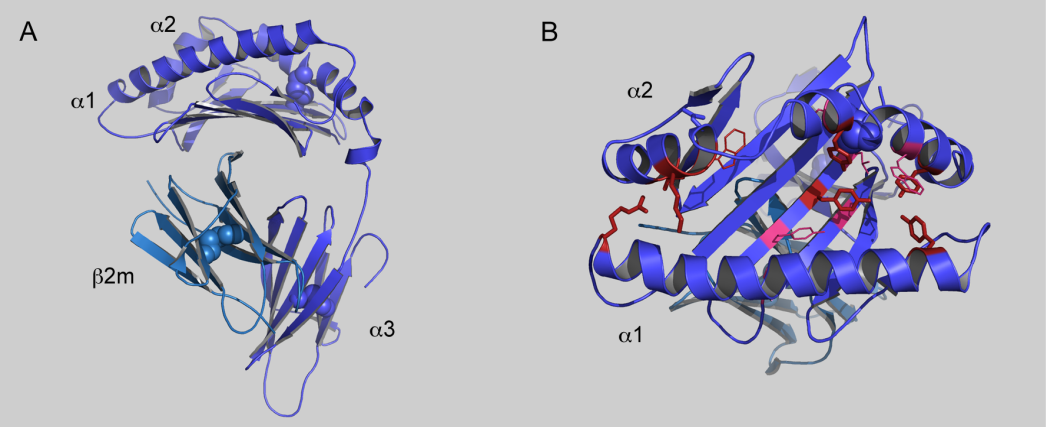


Figure 7. MHC-I structure. (A) MHC-I αβ heterodimer chains (Dila-UA/Dila-β2m predicted model), with α1, α2 and α3 regions and β2m-subunit indicated. (B) Top view of MHC-I heterodimer showing the peptide-binding region. Disulfide bonds as spheres. Side-chains of anchor residues close the ends of the groove, limiting the size of peptides that can bind. Transmembrane and cytoplasmic regions are not shown.

Also as referred above, MHC-II molecules form trimeric complexes that consist of non-covalently associated α and β chains and a loaded peptide. Each chain contains two extracellular domains (α1 and α2 or β1 and β2, respectively) (Fig. 8), a transmembrane and a cytoplasmic region [117]. Most of the polymorphism is found within the β1 domain of the β chain. The membrane-proximal α2 and β2 domains, both homologous with the Ig C1 domain, show very low polymorphism. The CD4 co-receptor (on αβ T cells) interacts with conserved region of β2 domain of class II. The membrane distal α1 and β1 domains

together constitute the platform of β-strands that supports two anti-parallel α-helices and form a peptide binding groove similar to that of MHC class I molecules (Fig. 8). However, the ends of the peptide-binding groove are open, allowing longer peptides of 12-25 or more amino acids to bind.

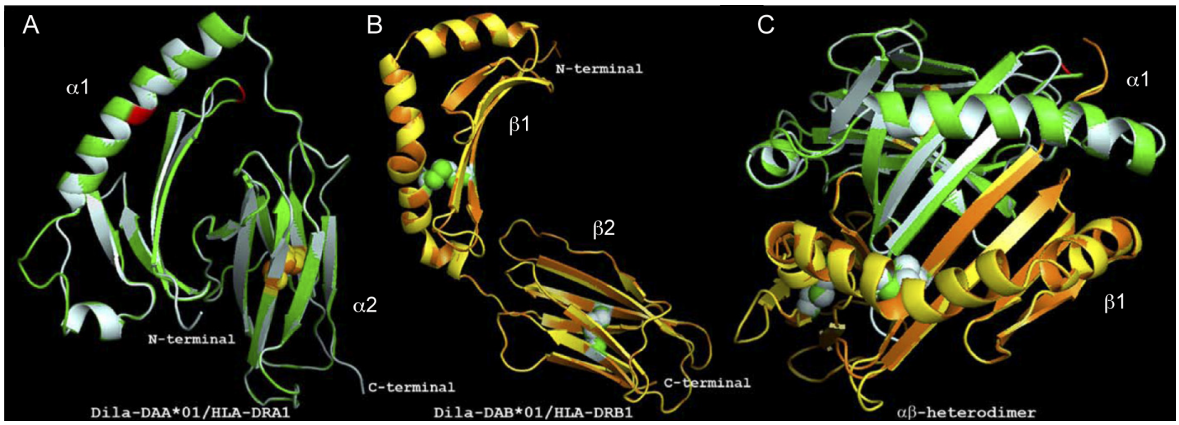


Figure 8. MHC-II structure. (A) MHC-II α chains (Dila-DAA/HLA-DRA1, green/white), with α1 and α2 regions indicated. (B) MHC-II β chains (Dila-DAB/HLADRB1, yellow/orange), with β1 and β2 domains denoted. (C) MHCII αβ-heterodimer viewed from top (peptide-binding region). Disulfide bonds are shown as space-filling model. Transmembrane and cytoplasmic regions are not shown. Adapted from [118].

3.6 MHC intracellular pathways

Peptide presentation is based on a complex system, involving protein degradation machinery, peptide transporters, chaperones for peptide loading and the MHC molecules themselves [75]. As mentioned above, there are two different antigen-processing pathways that serve to present peptides to TCRs on T lymphocytes. These pathways, and the machinery required for them, have distinct roles in the immune system, and function to sample different environments for antigenic peptides [119].

MHC-I molecules are loaded in the endoplasmic reticulum (ER) with peptides derived from degradation of cytosolic proteins by nuclear and cytosolic proteasomes, and these MHC-I/peptide complexes are then surface expressed displaying their antigenic cargo to cytotoxic to CD8+ T cells (Fig. 9), which can thereby efficiently recognize and eliminate infected or malignantly transformed cells. ER chaperone proteins such as calreticulin, ERp57, protein disulfide isomerase (PDI) and tapasin stabilize empty heterodimers. Tapasin interacts with TAP, coupling peptide translocation with delivery to class I molecules. Upon peptide binding, the chaperones are released and assembled MHC/peptide complexes leave the ER for presentation at the cell surface [74]. Peptides and class I molecules that fail to associate in the ER are returned to the cytosol for degradation [120].

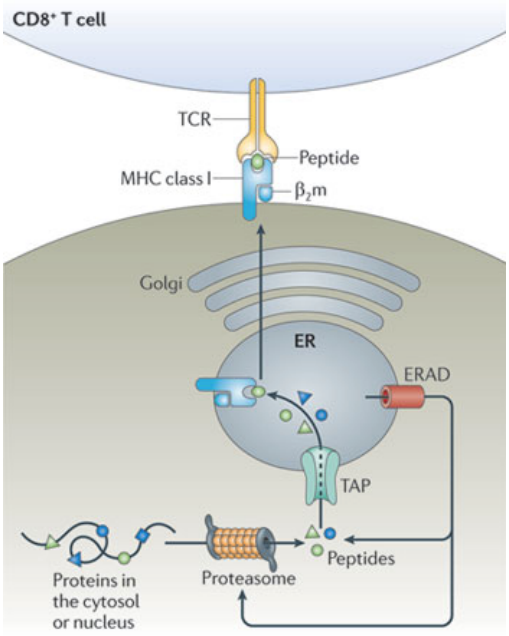


Figure 9. MHC-I antigen presentation pathway. The presentation of intracellular antigenic peptides by MHC-I molecules is a multi-step process. First, antigens are degraded by the proteasome. Then, the resulting peptides are translocated via transporter associated with antigen presentation (TAP) into the ER and loaded onto MHC-I molecules. MHC-I/peptide complexes are released from the ER and transported via the Golgi to the plasma membrane for antigen presentation to $CD8^+$ T cells. ERAD, ER-associated protein degradation. Adapted from [74].

MHC-II molecules exit the ER in association with the invariant chain occupying their peptide-binding groove. In the endocytic pathway proteolysis results in degradation of the invariant chain leaving a residual fragment (CLIP) in the binding groove [74]. With the help of HLA-DM CLIP is replaced by peptides resulting from the proteolysis of proteins resident or internalized into the endocytic pathway [74]. At the surface they are monitored by CD4+ T cells (Fig. 10).

Hence, MHC-I generally serves as a reporter of intracellular infection, while MHC-II senses the antigens present in the extracellular environment. Nevertheless, different mechanisms of cross-presentation of exogenous peptides by class I molecules [119] have been described, involving various cell types: endothelial [121] and other non-professional antigen presenting cells [122], and APCs, the most important being dendritic cells [123], potent activators of T cells. The different proposed routes of MHC I cross-presentation can be proteasome/TAP-dependent (cytosolic cross-presentation pathway) [124] or -independent (vacuolar cross-presentation pathway), requiring vacuolar cathepsins [125]. On the other hand, endogenous proteins can also be cross-presented by MHC class II molecules when these proteins are degraded through autophagy or other pathways [126].

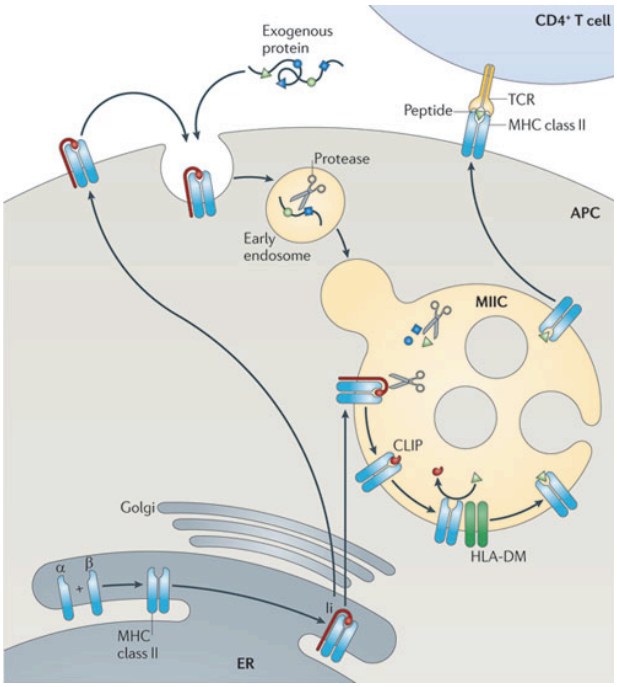


Figure 10. MHC-II antigen presentation pathway. MHC-II α- and β-chains assemble in the endoplasmic reticulum (ER) and form a complex with the invariant chain (Ii). The MHC-II/Ii heterotrimer is transported through the Golgi to the MHC class II compartment (MIIC), either directly and/or via the plasma membrane. Endocytosed proteins and Ii are degraded in the MIIC by resident proteases. The class II-associated Ii peptide (CLIP) fragment of Ii remains in the peptide-binding groove of the MHC-II dimer and is exchanged for an antigenic peptide with the help of the dedicated chaperone HLA-DM (known as H2-M in mice). MHC-II molecules are then transported to the plasma membrane to present antigenic peptides to CD4+ T cells. Adapted from [74].

3.7 Assembly of MHC-I molecules

Proper assembly of MHC-I/peptide complexes is required for their stable surface expression [127]. This process occurs in the ER and uses both conventional chaperones and dedicated factors. It is initiated in the cytosol by proteasomal degradation of proteins, originating small peptides [128]. The MHC-I peptidome includes DRiPs (defective ribosomal products), which are derived from non-functional proteins, and peptide fragments generated by ligation (non-encoded in the genome) [74]. The 26S proteasome generates peptides with defined C-termini, but extended at their N-termini. Two other proteasomes have been described: the immunoproteasome, expressed by immune cells; and the thymus-specific proteasome, expressed in thymic epithelial cells (TECs) [74], both altering the degradation pattern of the 26S proteasome [129, 130]. Immunoproteasomes are more active than 26S proteasomes under conditions of immune stress or IFNγ exposure [74]. Once generated, peptides of 8-16 residues are translocated by TAP into the ER, with ATP consumption [131]. ER aminopeptidases associated with antigen presentation (ERAAP/ERAP1 and ERAP2), can further trim the peptides within the ER. TAP also acts as a platform for folding MHC-I molecules by binding to one or more molecules of tapasin (the dedicated class I chaperone) [132]. In the ER, class I molecules are partially folded and stabilized by additional common chaperones: calnexin, calreticulin, ERp57, and peptide-binding chaperone PDI [133]. Together, calnexin and ERp57 catalyze MHC-I HC oxidative folding during early assembly events [134]. Subsequently, HC/β2m association occurs needing further stabilization. Calreticulin binds N-glycosylated MHC-I molecules, stabilizing the MHC-I HC/TPN interaction [135]. The complex of TAP, TPN,

MHC-I, ERp57 and CRT is called the peptide loading complex (PLC) (Fig. 11). The PLC ensures efficient peptide loading onto MHC-I molecules. PLC-associated PDI has been inconsistently found [134], Nevertheless, a role as a peptide carrier, protecting peptides from degradation and delivering them to the MHC in the PLC [136] and another maintaining HC oxidized state (re-oxidation of α2) [137] have been described. After binding to a peptide with sufficient affinity, the class I complex is released from the PLC and goes to the cell surface. Tapasin may also act as a peptide editor, resulting in loading of a more diverse array of peptides [138]. PDI has been further involved in PLC disassembly [139].

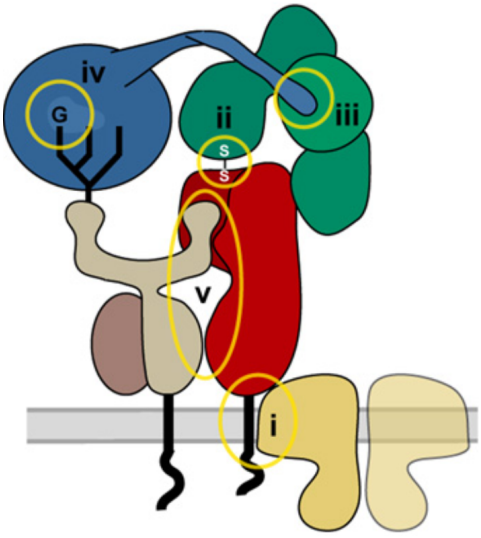


Figure 11. The PLC components and defined intermolecular interactions. (i) Tapasin (red) acts as a bridge between TAP (yellow) and the remaining PLC components. This interaction is important for stabilizing TAP [140] and the entire PLC assembly [141], and involves the TPN transmembrane (TM) domain and the N-terminal TM domains of TAP1 and TAP2 [141-143]. (ii) ERp57 (green) is covalently associated to TPN. This involves a disulfide bond between C^{95} of tapasin and C^{57} of ERp57, which is an active site cysteine [144]. Despite ERp57 typically forming transient mixed disulfides with its substrates, the conjugate is not a folding intermediate [145]. All TPN in the PLC is stably disulfide-linked to ERp57 [145, 146]. (iii) The tip of the CRT (blue) P-domain binds to residues in the b' domain of ERp57 [147]. (iv) CRT binds to a monoglucosylated N-linked glycan at N^{86} of the class I HC (brown) [148]. The glycan is sufficient to mediate HC binding [148] but the existence of additional polypeptide-based CRT interactions has been suggested [149]. (v) The interaction between the MHC-I HC and TPN includes: ER luminal domains of TPN interaction with two solvent-exposed regions in the ER luminal domains of HCs. One region comprises the α3 domain (residues 222–227/229), which is also the binding site for the CD8 co-receptor [150-152]. The second binding site is a loop in the α2 domain (residues 128–136), which lies adjacent to the carboxyl-terminal end of the peptide-binding groove and connects the short α2-1 helix to the β sheets underlying the peptide-binding groove [153-155]. Adapted from [132].

The MHC-I antigen-processing machinery can be affected in different ways by viruses, which can interfere either by transcriptional downregulation of antigen-processing machinery components [156], or by targeting the functionality of the PLC at the protein level [157]. Interestingly, host cells evolved strategies to potentially overcome the plethora of inhibitory mediators. Peptide translocation through gap junctions and cross-presentation by neighbouring cells would help limit the spread of infection, enabling pathogen-derived peptides from an infected cell that may have had its MHC machinery

disarmed to be presented [158]. It remains to be seen whether bacteria and other pathogens can also target the biosynthesis of MHC-I molecules [158]. Another way of circumventing viral immune evasion strategies that target the conventional pathway is by TAP-independent presentation mediated by autophagy, utilizing the vacuolar pathway [159]. Viral effects can also indirectly contribute to the inhibition of the IFN-mediated induction of MHC [160]. Likewise, many tumors also escape immune surveillance by irreversibly or reversibly downregulating MHC-I molecules [161]. Immune escape by contagious cancers through epigenetic downregulation of MHC molecules has been recently described [162].

4. Aim of this project

The European sea bass (*Dicentrarchus labrax)* was the first marine non-salmonid species to be commercially cultured in Europe and at present is the most important commercial fish cultured in Mediterranean areas (Food and Agriculture Organization, FAO). The annual world production of sea bass was estimated at 120,000 tons in 2010 (FAO), with Portugal among the world producers. Nevertheless, aquaculture remains an activity with a marginal contribution to the Portuguese national economy (FAO).

As species in rearing conditions are affected by a variety of infectious diseases caused by numerous pathogens, including viruses, bacteria, fungi and parasites, their economic importance to the aquaculture industry has been fuelling research on fish immunology. To prevent losses, it is mandatory to gain insights on the function and organisation of the immune system of such species, in order to improve health control.

The FIV laboratory at IBMC uses this species for studying host-pathogen interactions [163, 164]. Characterising the immune system of sea bass is an important goal, since a better understanding of the way in which it responds to infection will lead to improved vaccination strategies. On the other hand, being a bony fish offers a multitude of opportunities to explore the immune defences of this species.

As previously described, genes of the MHC have a crucial role in disease outcome, since they are one of the main players in adaptive immunity. This work aimed at gaining insights on sea bass MHC-I, namely through the identification and characterization at the genetic and structural levels of the class I complex subunits as well as the key elements of the processing pathway, for which there was no previous knowledge. The specific goals were:

- Clone, sequence and characterize the α and β subunits of sea bass MHC class I complexes
- Clone, sequence and characterize of the molecules involved in antigen processing pathway
- Use the obtained information to produce new tools, namely polyclonal and/or monoclonal antibodies, that could assist on further studies of the MHC class I presentation mechanism in this species, contributing to its understanding.

References

1. Loker ES. Macroevolutionary Immunology: A Role for Immunity in the Diversification of Animal life. Front Immunol. 2012 3:25.
2. Zhu LY, Nie L, Zhu G, Xiang LX, Shao JZ. Advances in research of fish immune-relevant genes: a comparative overview of innate and adaptive immunity in teleosts. Dev Comp Immunol. 2013 39:39-62.
3. Boehm T. Evolution of vertebrate immunity. Curr Biol. 2012 22:R722-32.
4. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. Nat Rev Genet. 2009 10:725-32.
5. Robert J. Comparative study of tumorigenesis and tumor immunity in invertebrates and nonmammalian vertebrates. Dev Comp Immunol. 2010 34:915-25.
6. Boehm T, Hess I, Swann JB. Evolution of lymphoid tissues. Trends Immunol. 2012 33:315-21.
7. Takeuchi O, Akira S. Pattern recognition receptors and inflammation. Cell. 2010 140:805-20.
8. Litman GW, Cooper MD. Why study the evolution of immunity? Nat Immunol. 2007 8:547-8.
9. Metchnikoff I. Lectures on the Comparative Pathology of Inflammation Delivered at the Pasteur Institute. Lectures on the Comparative Pathology of Inflammation Delivered at the Pasteur Institute; 1891.
10. Cooper MD, Peterson RD, Good RA. Delineation of the Thymic and Bursal Lymphoid Systems in the Chicken. Nature. 1965 205:143-6.
11. Reynaud CA, Anquez V, Grimal H, Weill JC. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell. 1987 48:379-88.
12. Hoffmann JA. The immune response of Drosophila. Nature. 2003 426:33-8.
13. Li J, Barreda DR, Zhang YA, Boshra H, Gelman AE, Lapatra S, et al. B lymphocytes from early vertebrates have potent phagocytic and microbicidal abilities. Nat Immunol. 2006 7:1116-24.
14. Litman GW, Rast JP, Fugmann SD. The origins of vertebrate adaptive immunity. Nat Rev Immunol. 2010 10:543-53.
15. Hirano M, Das S, Guo P, Cooper MD. The evolution of adaptive immunity in vertebrates. Adv Immunol. 2011 109:125-57.
16. Kasamatsu J. Evolution of innate and adaptive immune systems in jawless vertebrates. Microbiol Immunol. 2013 57:1-12.

17. Ghosh J, Lun CM, Majeske AJ, Sacchi S, Schrankel CS, Smith LC. Invertebrate immune diversity. Dev Comp Immunol. 2011 35:959-74.
18. Boehm T. Design principles of adaptive immune systems. Nat Rev Immunol. 2011 11:307-17.
19. Boehm T, McCurley N, Sutoh Y, Schorpp M, Kasahara M, Cooper MD. VLR-based adaptive immunity. Annu Rev Immunol. 2012 30:203-20.
20. Kurtz J, Franz K. Innate defence: evidence for memory in invertebrate immunity. Nature. 2003 425:37-8.
21. Kurtz J. Specific memory within innate immune systems. Trends Immunol. 2005 26:186-92.
22. Sadd BM, Schmid-Hempel P. Insect immunity shows specificity in protection upon secondary pathogen exposure. Curr Biol. 2006 16:1206-10.
23. Netea MG, Quintin J, van der Meer JW. Trained immunity: a memory for innate host defense. Cell Host Microbe. 2011 9:355-61.
24. Silverstein AM. The Clonal Selection Theory: what it really is and why modern challenges are misplaced. Nat Immunol. 2002 3:793-6.
25. Paust S, von Andrian UH. Natural killer cell memory. Nat Immunol. 2011 12:500-8.
26. Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz RA. Phylogenetic perspectives in innate immunity. Science. 1999 284:1313-8.
27. Iwasaki A, Medzhitov R. Toll-like receptor control of the adaptive immune responses. Nat Immunol. 2004 5:987-95.
28. Zhang SM, Adema CM, Kepler TB, Loker ES. Diversification of Ig superfamily genes in an invertebrate. Science. 2004 305:251-4.
29. Watson FL, Puttmann-Holgado R, Thomas F, Lamar DL, Hughes M, Kondo M, et al. Extensive diversity of Ig-superfamily proteins in the immune system of insects. Science. 2005 309:1874-8.
30. Laird DJ, De Tomaso AW, Weissman IL. Stem cells are units of natural selection in a colonial ascidian. Cell. 2005 123:1351-60.
31. Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet. 2010 11:47-59.
32. Boehm T, Iwanami N, Hess I. Evolution of the immune system in the lower vertebrates. Annu Rev Genomics Hum Genet. 2012 13:127-49.
33. Boehm T. Co-evolution of a primordial peptide-presentation system and cellular immunity. Nat Rev Immunol. 2006 6:79-84.
34. McCurley N, Hirano M, Das S, Cooper MD. Immune related genes underpin the evolution of adaptive immunity in jawless vertebrates. Curr Genomics. 2012 13:86-

94.

35. Guo P, Hirano M, Herrin BR, Li J, Yu C, Sadlonova A, et al. Dual nature of the adaptive immune system in lampreys. Nature. 2009 459:796-801.
36. Alder MN, Herrin BR, Sadlonova A, Stockard CR, Grizzle WE, Gartland LA, et al. Antibody responses of variable lymphocyte receptors in the lamprey. Nat Immunol. 2008 9:319-27.
37. Kasamatsu J, Sutoh Y, Fugo K, Otsuka N, Iwabuchi K, Kasahara M. Identification of a third variable lymphocyte receptor in the lamprey. Proc Natl Acad Sci U S A. 2010 107:14304-8.
38. Kishishita N, Matsuno T, Takahashi Y, Takaba H, Nishizumi H, Nagawa F. Regulation of antigen-receptor gene assembly in hagfish. EMBO Rep. 2010 11:126-32.
39. Pancer Z, Amemiya CT, Ehrhardt GR, Ceitlin J, Gartland GL, Cooper MD. Somatic diversification of variable lymphocyte receptors in the agnathan sea lamprey. Nature. 2004 430:174-80.
40. Nagawa F, Kishishita N, Shimizu K, Hirose S, Miyoshi M, Nezu J, et al. Antigen-receptor genes of the agnathan lamprey are assembled by a process involving copy choice. Nat Immunol. 2007 8:206-13.
41. Rogozin IB, Iyer LM, Liang L, Glazko GV, Liston VG, Pavlov YI, et al. Evolution and diversification of lamprey antigen receptors: evidence for involvement of an AID-APOBEC family cytosine deaminase. Nat Immunol. 2007 8:647-56.
42. Pancer Z, Cooper MD. The evolution of adaptive immunity. Annu Rev Immunol. 2006 24:497-518.
43. Reynaud CA, Bertocci B, Dahan A, Weill JC. Formation of the chicken B-cell repertoire: ontogenesis, regulation of Ig gene rearrangement, and diversification by gene conversion. Adv Immunol. 1994 57:353-78.
44. Kreslavsky T, Gleimer M, von Boehmer H. Alphabeta versus gammadelta lineage choice at the first TCR-controlled checkpoint. Curr Opin Immunol. 2010 22:185-92.
45. Chaudhuri J, Basu U, Zarrin A, Yan C, Franco S, Perlot T, et al. Evolution of the immunoglobulin heavy chain class switch recombination mechanism. Adv Immunol. 2007 94:157-214.
46. Iwasaki A, Medzhitov R. Regulation of adaptive immunity by the innate immune system. Science. 2010 327:291-5.
47. Geissmann F, Manz MG, Jung S, Sieweke MH, Merad M, Ley K. Development of monocytes, macrophages, and dendritic cells. Science. 2010 327:656-61.

48. Aghaallaei N, Bajoghli B, Schwarz H, Schorpp M, Boehm T. Characterization of mononuclear phagocytic cells in medaka fish transgenic for a cxcr3a:gfp reporter. Proc Natl Acad Sci U S A. 2010 107:18079-84.
49. Lugo-Villarino G, Balla KM, Stachura DL, Banuelos K, Werneck MB, Traver D. Identification of dendritic antigen-presenting cells in the zebrafish. Proc Natl Acad Sci U S A. 2010 107:15850-5.
50. Haugland GT, Jordal AE, Wergeland HI. Characterization of small, mononuclear blood cells from salmon having high phagocytic capacity and ability to differentiate into dendritic like cells. PLoS One. 2012 7:e49260.
51. Bassity E, Clark TG. Functional identification of dendritic cells in the teleost model, rainbow trout (Oncorhynchus mykiss). PLoS One. 2012 7:e33196.
52. Rumfelt LL, McKinney EC, Taylor E, Flajnik MF. The development of primary and secondary lymphoid tissues in the nurse shark Ginglymostoma cirratum: B-cell zones precede dendritic cell immigration and T-cell zone formation during ontogeny of the spleen. Scand J Immunol. 2002 56:130-48.
53. Ohta Y, Landis E, Boulay T, Phillips RB, Collet B, Secombes CJ, et al. Homologs of CD83 from elasmobranch and teleost fish. J Immunol. 2004 173:4553-60.
54. Boehm T, Bleul CC. The evolutionary history of lymphoid organs. Nat Immunol. 2007 8:131-5.
55. Pickel JM, McCormack WT, Chen CH, Cooper MD, Thompson CB. Differential regulation of V(D)J recombination during development of avian B and T cells. Int Immunol. 1993 5:919-27.
56. Brendolan A, Rosado MM, Carsetti R, Selleri L, Dear TN. Development and function of the mammalian spleen. Bioessays. 2007 29:166-77.
57. Boehm T, Bleul CC. Thymus-homing precursors and the thymic microenvironment. Trends Immunol. 2006 27:477-84.
58. Bhandoola A, von Boehmer H, Petrie HT, Zuniga-Pflucker JC. Commitment and developmental potential of extrathymic and intrathymic T cell precursors: plenty to choose from. Immunity. 2007 26:678-89.
59. Anderson G, Lane PJ, Jenkinson EJ. Generating intrathymic microenvironments to establish T-cell tolerance. Nat Rev Immunol. 2007 7:954-63.
60. Bajoghli B, Guo P, Aghaallaei N, Hirano M, Strohmeier C, McCurley N, et al. A thymus candidate in lampreys. Nature. 2011 470:90-4.
61. Magnadottir B. Innate immunity of fish (overview). Fish Shellfish Immunol. 2006 20:137-51.
62. Magor BG, Magor KE. Evolution of effectors and receptors of innate immunity. Dev

Comp Immunol. 2001 25:651-82.

63. Vasta GR, Nita-Lazar M, Giomarelli B, Ahmed H, Du S, Cammarata M, et al. Structural and functional diversity of the lectin repertoire in teleost fish: relevance to innate and adaptive immunity. Dev Comp Immunol. 2011 35:1388-99.
64. Yoder JA, Litman GW. The phylogenetic origins of natural killer receptors and recognition: relationships, possibilities, and realities. Immunogenetics. 2011 63:123-41.
65. Sunyer JO, Zarkadis IK, Lambris JD. Complement diversity: a mechanism for generating immune diversity? Immunol Today. 1998 19:519-23.
66. Zapata A, Amemiya CT. Phylogeny of lower vertebrates and their immunological structures. Curr Top Microbiol Immunol. 2000 248:67-107.
67. Salinas I, Zhang YA, Sunyer JO. Mucosal immunoglobulins and B cells of teleost fish. Dev Comp Immunol. 2011 35:1346-65.
68. Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat Immunol. 2005 6:295-302.
69. Hansen JD, Landis ED, Phillips RB. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish. Proc Natl Acad Sci U S A. 2005 102:6919-24.
70. Zhang YA, Salinas I, Li J, Parra D, Bjork S, Xu Z, et al. IgT, a primitive immunoglobulin class specialized in mucosal immunity. Nat Immunol. 2010 11:827-35.
71. Laing KJ, Hansen JD. Fish T cells: recent advances through genomics. Dev Comp Immunol. 2011 35:1282-95.
72. Sunyer JO. Fishing for mammalian paradigms in the teleost immune system. Nat Immunol. 2013 14:320-6.
73. Rauta PR, Nayak B, Das S. Immune system and immune responses in fish and their role in comparative immunity study: a model for higher organisms. Immunol Lett. 2012 148:23-33.
74. Neefjes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol. 2011 11:823-36.
75. Trombetta ES, Mellman I. Cell biology of antigen processing in vitro and in vivo. Annu Rev Immunol. 2005 23:975-1028.
76. Milinski M, Griffiths S, Wegner KM, Reusch TB, Haas-Assenbaum A, Boehm T. Mate choice decisions of stickleback females predictably modified by MHC peptide ligands. Proc Natl Acad Sci U S A. 2005 102:4414-8.

77. Spehr M, Kelliher KR, Li XH, Boehm T, Leinders-Zufall T, Zufall F. Essential role of the main olfactory system in social recognition of major histocompatibility complex peptide ligands. J Neurosci. 2006 26:1961-70.
78. Leinders-Zufall T, Ishii T, Mombaerts P, Zufall F, Boehm T. Structural requirements for the activation of vomeronasal sensory neurons by MHC peptides. Nat Neurosci. 2009 12:1551-8.
79. Leinders-Zufall T, Brennan P, Widmayer P, S PC, Maul-Pavicic A, Jager M, et al. MHC class I peptides as chemosensory signals in the vomeronasal organ. Science. 2004 306:1033-7.
80. Milinski M, Croy I, Hummel T, Boehm T. Major histocompatibility complex peptide ligands as olfactory cues in human body odour assessment. Proc Biol Sci. 2013 280:20122889.
81. Little CC, Tyzzer EE. Further experimental studies on the inheritance of susceptibility to a Transplantable tumor, Carcinoma (J. W. A.) of the Japanese waltzing Mouse. J Med Res. 1916 33:393-453.
82. Gorer PA. The genetic and antigenic basis of tumour transplantation. J Pathol Bacteriol. 1937 44:691-7.
83. Snell GD. Methods for the study of histocompatibility genes. J Genet. 1948 49:87-108.
84. Flajnik MF, Kasahara M. Comparative genomics of the MHC: glimpses into the evolution of the adaptive immune system. Immunity. 2001 15:351-62.
85. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet. 2009 54:15-39.
86. Horton R, Wilming L, Rand V, Lovering RC, Bruford EA, Khodiyar VK, et al. Gene map of the extended human MHC. Nat Rev Genet. 2004 5:889-99.
87. Kelley J, Walter L, Trowsdale J. Comparative genomics of major histocompatibility complexes. Immunogenetics. 2005 56:683-95.
88. Ohta Y, Shiina T, Lohr RL, Hosomichi K, Pollin TI, Heist EJ, et al. Primordial linkage of beta2-microglobulin to the MHC. J Immunol. 2011 186:3563-71.
89. Deng L, Mariuzza RA. Structural basis for recognition of MHC and MHC-like ligands by natural killer cell receptors. Semin Immunol. 2006 18:159-66.
90. Parham P, Moffett A. Variable NK cell receptors and their MHC class I ligands in immunity, reproduction and human evolution. Nat Rev Immunol. 2013 13:133-44.
91. Stet RJ, Kruiswijk CP, Dixon B. Major histocompatibility lineages and immune gene function in teleost fishes: the road not taken. Crit Rev Immunol. 2003 23:441-71.
92. Kaufman J, Milne S, Gobel TW, Walker BA, Jacob JP, Auffray C, et al. The chicken

B locus is a minimal essential major histocompatibility complex. Nature. 1999 401:923-5.

93. Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF. Ancestral organization of the MHC revealed in the amphibian Xenopus. J Immunol. 2006 176:3674-85.
94. Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF. Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc Natl Acad Sci U S A. 2000 97:4712-7.
95. Neefjes JJ, Ploegh HL. Allele and locus-specific differences in cell surface expression and the association of HLA class I heavy chain with beta 2-microglobulin: differential effects of inhibition of glycosylation on class I subunit association. Eur J Immunol. 1988 18:801-10.
96. Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, et al. Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. J Immunol. 2008 181:3393-9.
97. Grimholt U, Drablos F, Jorgensen SM, Hoyheim B, Stet RJ. The major histocompatibility class I locus in Atlantic salmon (Salmo salar L.): polymorphism, linkage analysis and protein modelling. Immunogenetics. 2002 54:570-81.
98. Shum BP, Guethlein L, Flodin LR, Adkison MA, Hedrick RP, Nehring RB, et al. Modes of salmonid MHC class I and II evolution differ from the primate paradigm. J Immunol. 2001 166:3297-308.
99. Aoyagi K, Dijkstra JM, Xia C, Denda I, Ototake M, Hashimoto K, et al. Classical MHC class I genes composed of highly divergent sequence lineages share a single locus in rainbow trout (Oncorhynchus mykiss). J Immunol. 2002 168:260-73.
100. Miller KM, Kaukinen KH, Schulze AD. Expansion and contraction of major histocompatibility complex genes: a teleostean example. Immunogenetics. 2002 53:941-63.
101. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, et al. The genome sequence of Atlantic cod reveals a unique immune system. Nature. 2011 477:207-10.
102. Tsukamoto K, Hayashi S, Matsuo MY, Nonaka MI, Kondo M, Shima A, et al. Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, Oryzias latipes. Immunogenetics. 2005 57:420-31.
103. Sato A, Figueroa F, O'HUigin C, Steck N, Klein J. Cloning of major histocompatibility complex (Mhc) genes from threespine stickleback, Gasterosteus aculeatus. Mol Mar Biol Biotechnol. 1998 7:221-31.
104. Kaufman J. Antigen processing and presentation: Evolution from a bird's eye view. Mol Immunol. 2013 55:159-61.

105. Ohta Y, Powis SJ, Lohr RL, Nonaka M, Pasquier LD, Flajnik MF. Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in Xenopus: further evidence for co-evolution among MHC class I region genes. Eur J Immunol. 2003 33:3017-27.
106. Dijkstra JM, Katagiri T, Hosomichi K, Yanagiya K, Inoko H, Ototake M, et al. A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. Immunogenetics. 2007 59:305-21.
107. Lukacs MF, Harstad H, Bakke HG, Beetz-Sargent M, McKinnel L, Lubieniecki KP, et al. Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. BMC Genomics. 2010 11:154.
108. Stet RJ, Kruiswijk CP, Saeij JP, Wiegertjes GF. Major histocompatibility genes in cyprinid fishes: theory and practice. Immunol Rev. 1998 166:301-16.
109. Nei M, Rooney AP. Concerted and birth-and-death evolution of multigene families. Annu Rev Genet. 2005 39:121-52.
110. Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. Proc Natl Acad Sci U S A. 1997 94:7799-806.
111. Wegner KM. Historical and contemporary selection of teleost MHC genes: did we leave the past behind? Journal of Fish Biology. 2008 73:2110-32.
112. Mehta RB, Nonaka MI, Nonaka M. Comparative genomic analysis of the major histocompatibility complex class I region in the teleost genus Oryzias. Immunogenetics. 2009 61:385-99.
113. Eizaguirre C, Lenz TL, Kalbe M, Milinski M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat Commun. 2012 3:621.
114. Piertney SB, Oliver MK. The evolutionary ecology of the major histocompatibility complex. Heredity (Edinb). 2006 96:7-21.
115. van Oosterhout C. A new theory of MHC evolution: beyond selection on the immune genes. Proc Biol Sci. 2009 276:657-65.
116. Edwards SV, Hedrick PW. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol Evol. 1998 13:305-11.
117. Madden DR. The three-dimensional structure of peptide-MHC complexes. Annu Rev Immunol. 1995 13:587-622.
118. Silva DS, Reis MI, Nascimento DS, do Vale A, Pereira PJ, dos Santos NM. Sea bass (Dicentrarchus labrax) invariant chain and class II major histocompatibility complex: sequencing and structural analysis using 3D homology modelling. Mol

Immunol. 2007 44:3758-76.

119. Wagner CS, Grotzke JE, Cresswell P. Intracellular events regulating cross-presentation. Front Immunol. 2012 3:138.
120. Hughes EA, Hammond C, Cresswell P. Misfolded major histocompatibility complex class I heavy chains are translocated into the cytoplasm and degraded by the proteasome. Proc Natl Acad Sci U S A. 1997 94:1896-901.
121. Bagai R, Valujskikh A, Canaday DH, Bailey E, Lalli PN, Harding CV, et al. Mouse endothelial cells cross-present lymphocyte-derived antigen on class I MHC via a TAP1- and proteasome-dependent pathway. J Immunol. 2005 174:7711-5.
122. Giodini A, Rahner C, Cresswell P. Receptor-mediated phagocytosis elicits cross-presentation in nonprofessional antigen-presenting cells. Proc Natl Acad Sci U S A. 2009 106:3324-9.
123. Jung S, Unutmaz D, Wong P, Sano G, De los Santos K, Sparwasser T, et al. In vivo depletion of CD11c+ dendritic cells abrogates priming of CD8+ T cells by exogenous cell-associated antigens. Immunity. 2002 17:211-20.
124. Amigorena S, Savina A. Intracellular mechanisms of antigen cross presentation in dendritic cells. Curr Opin Immunol. 2010 22:109-17.
125. Tiwari N, Garbi N, Reinheckel T, Moldenhauer G, Hammerling GJ, Momburg F. A transporter associated with antigen-processing independent vacuolar pathway for the MHC class I-mediated presentation of endogenous transmembrane proteins. J Immunol. 2007 178:7932-42.
126. Crotzer VL, Blum JS. Autophagy and adaptive immunity. Immunology. 2010 131:9-17.
127. Pamer E, Cresswell P. Mechanisms of MHC class I--restricted antigen processing. Annu Rev Immunol. 1998 16:323-58.
128. Sijts EJ, Kloetzel PM. The role of the proteasome in the generation of MHC class I ligands and immune responses. Cell Mol Life Sci. 2011 68:1491-502.
129. Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. J Exp Med. 2001 194:1-12.
130. Nitta T, Murata S, Sasaki K, Fujii H, Ripen AM, Ishimaru N, et al. Thymoproteasome shapes immunocompetent repertoire of CD8+ T cells. Immunity. 2010 32:29-40.
131. Abele R, Tampé R. Peptide trafficking and translocation across membranes in cellular signaling and self-defense strategies. Current Opinion in Cell Biology. 2009 21:508-15.
132. Wearsch PA, Cresswell P. The quality control of MHC class I peptide loading. Curr Opin Cell Biol. 2008 20:624-31.

133. Kim Y, Kang K, Kim I, Lee YJ, Oh C, Ryoo J, et al. Molecular mechanisms of MHC class I-antigen processing: redox considerations. Antioxid Redox Signal. 2009 11:907-36.
134. Van Hateren A, James E, Bailey A, Phillips A, Dalchau N, Elliott T. The cell biology of major histocompatibility complex class I assembly: towards a molecular understanding. Tissue Antigens. 2010 76:259-75.
135. Wearsch PA, Peaper DR, Cresswell P. Essential glycan-dependent interactions optimize MHC class I peptide loading. Proc Natl Acad Sci U S A. 2011 108:4950-5.
136. Cho K, Cho S, Lee SO, Oh C, Kang K, Ryoo J, et al. Redox-regulated peptide transfer from the transporter associated with antigen processing to major histocompatibility complex class I molecules by protein disulfide isomerase. Antioxid Redox Signal. 2011 15:621-33.
137. Park B, Lee S, Kim E, Cho K, Riddell SR, Cho S, et al. Redox regulation facilitates optimal peptide selection by MHC class I during antigen processing. Cell. 2006 127:369-82.
138. Zarling AL, Luckey CJ, Marto JA, White FM, Brame CJ, Evans AM, et al. Tapasin is a facilitator, not an editor, of class I MHC peptide binding. J Immunol. 2003 171:5287-95.
139. Lee S, Park B, Kang K, Ahn K. Redox-regulated export of the major histocompatibility complex class I-peptide complexes from the endoplasmic reticulum. Mol Biol Cell. 2009 20:3285-94.
140. Garbi N, Tiwari N, Momburg F, Hammerling GJ. A major role for tapasin as a stabilizer of the TAP peptide transporter and consequences for MHC class I expression. Eur J Immunol. 2003 33:264-73.
141. Leonhardt RM, Keusekotten K, Bekpen C, Knittler MR. Critical role for the tapasin-docking site of TAP2 in the functional integrity of the MHC class I-peptide-loading complex. J Immunol. 2005 175:5104-14.
142. Procko E, Raghuraman G, Wiley DC, Raghavan M, Gaudet R. Identification of domain boundaries within the N-termini of TAP1 and TAP2 and their importance in tapasin binding and tapasin-mediated increase in peptide loading of MHC class I. Immunol Cell Biol. 2005 83:475-82.
143. Koch J, Guntrum R, Tampe R. The first N-terminal transmembrane helix of each subunit of the antigenic peptide transporter TAP is essential for independent tapasin binding. FEBS Lett. 2006 580:4091-6.
144. Dick TP, Bangia N, Peaper DR, Cresswell P. Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. Immunity. 2002 16:87-98.

145. Peaper DR, Wearsch PA, Cresswell P. Tapasin and ERp57 form a stable disulfide-linked dimer within the MHC class I peptide-loading complex. EMBO J. 2005 24:3613-23.
146. Peaper DR, Cresswell P. The redox activity of ERp57 is not essential for its functions in MHC class I peptide loading. Proc Natl Acad Sci U S A. 2008 105:10477-82.
147. Frickel EM, Riek R, Jelesarov I, Helenius A, Wuthrich K, Ellgaard L. TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. Proc Natl Acad Sci U S A. 2002 99:1954-9.
148. Wearsch PA, Jakob CA, Vallin A, Dwek RA, Rudd PM, Cresswell P. Major histocompatibility complex class I molecules expressed with monoglucosylated N-linked glycans bind calreticulin independently of their assembly status. J Biol Chem. 2004 279:25112-21.
149. Ireland BS, Brockmeier U, Howe CM, Elliott T, Williams DB. Lectin-deficient calreticulin retains full functionality as a chaperone for class I histocompatibility molecules. Mol Biol Cell. 2008 19:2413-23.
150. Kulig K, Nandi D, Bacik I, Monaco JJ, Vukmanovic S. Physical and functional association of the major histocompatibility complex class I heavy chain alpha3 domain with the transporter associated with antigen processing. J Exp Med. 1998 187:865-74.
151. Suh WK, Derby MA, Cohen-Doyle MF, Schoenhals GJ, Fruh K, Berzofsky JA, et al. Interaction of murine MHC class I molecules with tapasin and TAP enhances peptide loading and involves the heavy chain alpha3 domain. J Immunol. 1999 162:1530-40.
152. Harris MR, Lybarger L, Myers NB, Hilbert C, Solheim JC, Hansen TH, et al. Interactions of HLA-B27 with the peptide loading complex as revealed by heavy chain mutations. Int Immunol. 2001 13:1275-82.
153. Lewis JW, Neisig A, Neefjes J, Elliott T. Point mutations in the alpha 2 domain of HLA-A2.1 define a functionally relevant interaction with TAP. Curr Biol. 1996 6:873-83.
154. Peace-Brewer AL, Tussey LG, Matsui M, Li G, Quinn DG, Frelinger JA. A point mutation in HLA-A*0201 results in failure to bind the TAP complex and to present virus-derived peptides to CTL. Immunity. 1996 4:505-14.
155. Yu YY, Turnquist HR, Myers NB, Balendiran GK, Hansen TH, Solheim JC. An extensive region of an MHC class I alpha 2 domain loop influences interaction with the assembly complex. J Immunol. 1999 163:4427-33.

156. Loch S, Tampe R. Viral evasion of the MHC class I antigen-processing machinery. Pflugers Arch. 2005 451:409-17.
157. Roder G, Geironson L, Bressendorff I, Paulsson K. Viral proteins interfering with antigen presentation target the major histocompatibility complex class I peptide-loading complex. J Virol. 2008 82:8246-52.
158. Antoniou AN, Powis SJ. Pathogen evasion strategies for the major histocompatibility complex class I assembly pathway. Immunology. 2008 124:1-12.
159. Tey SK, Khanna R. Autophagy mediates transporter associated with antigen processing-independent presentation of viral epitopes through MHC class I pathway. Blood. 2012 120:994-1004.
160. Samuel CE. Antiviral actions of interferons. Clin Microbiol Rev. 2001 14:778-809, table of contents.
161. Bubenik J. MHC class I down-regulation: tumour escape from immune surveillance? (review). Int J Oncol. 2004 25:487-91.
162. Siddle HV, Kreiss A, Tovar C, Yuen CK, Cheng Y, Belov K, et al. Reversible epigenetic down-regulation of MHC molecules by devil facial tumour disease illustrates immune escape by a contagious cancer. Proc Natl Acad Sci U S A. 2013 110:5103-8.
163. do Vale A, Costa-Ramos C, Silva A, Silva DS, Gartner F, dos Santos NM, et al. Systemic macrophage and neutrophil destruction by secondary necrosis induced by a bacterial exotoxin in a Gram-negative septicaemia. Cell Microbiol. 2007 9:988-1003.
164. do Vale A, Silva MT, dos Santos NM, Nascimento DS, Reis-Rodrigues P, Costa-Ramos C, et al. AIP56, a novel plasmid-encoded virulence factor of Photobacterium damselae subsp. piscicida with apoptogenic activity against sea bass macrophages and neutrophils. Mol Microbiol. 2005 58:1025-38.

CHAPTER 2

Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) MHC class I heavy chain and β2-microglobulin

Rute D. Pinto, Elisa Randelli, Francesco Buonocoro, Pedro J. B. Pereira and Nuno M. S. dos Santos

Dev Comp Immunol. 2013, 39 (3): 234-254, doi:10.1016/j.dci.2012.10.002

Developmental and Comparative Immunology xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) MHC class I heavy chain and β2-microglobulin

Rute D. Pinto [a,c], Elisa Randelli [d], Francesco Buonocore [d], Pedro J.B. Pereira [b], Nuno M.S. dos Santos [a,*]

[a] Fish Immunology and Vaccinology Group, Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal
[b] Biomolecular Structure Group, Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal
[c] Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Rua de Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal
[d] Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), University of Tuscia, Largo dell'Università snc, 01100 Viterbo, Italy

ARTICLE INFO

Article history:
Received 26 April 2011
Revised 10 October 2012
Accepted 11 October 2012
Available online xxxx

Keywords:
Dicentrarchus labrax
β2m
MHC class I
Gene cloning
Tissue-specific expression
Homology modeling

ABSTRACT

In this work, the gene and cDNA of sea bass (*Dicentrarchus labrax*) β2-microglobulin (*Dila-β2m*) and several cDNAs of MHC class I heavy chain (Dila-UA) were characterized. While Dila-β2m is single-copy, numerous Dila-UA transcripts were identified per individual with variability at the peptide-binding domain (PBD), but also with unexpected diversity from the connective peptide (CP) through the 3′ untranslated region (UTR). Phylogenetic analysis segregates Dila-β2m and Dila-UA into each subfamily cluster, placing them in the fish class and branching Dila-MHC-I with lineage U. The α1 domains resemble those of the recently proposed L1 trans-species lineage. Although no Dila-specific α1, α2 or α3 sub-lineages could be observed, two highly distinct sub-lineages were identified at the CP/TM/CYT regions. The three-dimensional homology model of sea bass MHC-I complex is consistent with other characterized vertebrate structures. Furthermore, basal tissue-specific expression profiles were determined for both molecules, and expression of β2m was evaluated after poly I:C stimulus. Results suggest these molecules are orthologues of other β2m and teleost classical MHC-I and their basic structure is evolutionarily conserved, providing relevant information for further studies on antigen presentation in this fish species.

1. Introduction

Class I major histocompatibility complexes (MHC) are heterotrimers formed by a membrane-anchored α-chain, a soluble β2-microglobulin (β2m) moiety and an antigen (Bjorkman et al., 1987b). MHC genes have a crucial role in the adaptive immunity of jawed vertebrates (Flajnik and Kasahara, 2010). Classical class I genes are polymorphic, exist on nearly all cell types, and present either self or foreign peptides from proteasomal degradation of cytosolic proteins to T cell receptors (TCRs) on cytotoxic $CD8^+$ T cells. Once activated by presentation of an altered/non-selfpeptide, the $CD8^+$ T cell kills the peptide-presenting cell (Rudolph et al., 2006). β2-Microglobulin is not covalently attached to the class I heavy chain (HC), and is freely exchangeable with blood β2m *in vivo* (Kimura et al., 1983). Besides classical MHC or class Ia antigens (Grey et al., 1973), β2m also associates with most non-classical or class Ib molecules (including HLA-E, -F, -G, and -H) and other MHC-related proteins (such as CD1 (Sugita et al., 1997) and FcRn (Simister and Ahouse, 1996)). Assembly of the trimeric complexes is a tightly regulated process (Peaper and Cresswell, 2008) occurring in the endoplasmic reticulum (ER) with assistance of several chaperones that form the peptide loading complex (PLC), where β2m is essential for proper folding, peptide binding and surface expression of the assembled molecules (Hansen et al., 1988; Vitiello et al., 1990). The PLC, which stabilizes the empty HC-β2m dimer and helps loading high affinity peptides (Diedrich et al., 2001; Wearsch and Cresswell, 2008), includes calreticulin (CRT), ER protein of 57 kDa (ERp57), tapasin (TPN) and the transporter-associated with antigen processing (TAP). The TAP1/TAP2 heterodimer transports peptides from the cytosol into the ER, while TPN makes the bridge between the class I itself and the transporter (Sadasivan et al., 1996), being important for peptide-editing (Chen and Bouvier, 2007; Praveen et al., 2010). Tapasin also covalently binds to ERp57 (Peaper et al., 2005) forming the conjugate considered to be the functional unit of the PLC (Wearsch and Cresswell, 2007), thereby protecting the empty peptide-binding groove against α2 disulfide reduction, until that protection is provided by a peptide ligand (Kienast et al., 2007). Other interactions between the PLC components include the contact between CRT and ERp57 (Frickel et al., 2002), and the lectin-like association of CRT with the monoglycosylated N-linked glycan on the HC (Sadasivan et al., 1996; Wearsch et al., 2004).

* Corresponding author. Tel.: +351 226 074 900; fax: +351 226 099 157.
E-mail addresses: rsp@ibmc.up.pt (R.D. Pinto), elisa.randelli@unitus.it (E. Randelli), fbuono@unitus.it (F. Buonocore), ppereira@ibmc.up.pt (P.J.B. Pereira), nsantos@ibmc.up.pt (N.M.S. dos Santos).

0145-305X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.dci.2012.10.002

The three-dimensional structures of bovine, human and grass carp β2m have been elucidated (Becker and Reeke, 1985; Bjorkman et al., 1987b; Chen et al., 2010b), unveiling an immunoglobulin-like domain, with a seven-stranded anti-parallel β-sandwich stabilized by a single disulfide bond (Trinh et al., 2002). Also the crystal structures of human (Bjorkmann et al., 1987; Saper et al., 1991), mouse (Achour et al., 1998), chicken (Koch et al., 2007) and bovine (Macdonald et al., 2010) MHC class I trimeric complexes (extracellular HC domains, β2m and epitope) have been determined. The MHC class I heavy chain typically consists of a peptide-binding domain (PBD) formed by two adjacent helical regions (α1 and α2) arranged on top of an eight-stranded β-sheet. The antigen-binding cleft, containing six pockets for peptide side-chain binding, is characteristically tapered and blocked at both ends by bulky side-chains. The side-chains of the most highly variable residues of the PBD point into the groove and up from the top of both helices, while highly conserved residues are found at both ends of the groove (Madden, 1995). The peptide is held in an extended center-arched or zigzag conformation (Batalia and Collins, 1997), running along the length of the cleft. Below the PBD there is also a typical Ig-C folding domain (α3 region), making the connection to the transmembrane anchor and cytoplasmic domain (Madden, 1995). While the position of the α3 domain relative to the PBD seems to be variable, that of the β2m is well conserved due to extensive binding to both α1 and α2 (Koch et al., 2007; Macdonald et al., 2010).

Since the characterization of human β2m and HC in the 80s (Bjorkman et al., 1987b; Gussow et al., 1987), both molecules have been identified in many jawed vertebrates, from mammals to bony and cartilaginous fish. A distinct genomic organization of the MHC region has been described for bony fish: unlike all other jawed vertebrates, classical class I and II genes are not linked (Bingulac-Popovic et al., 1997; Hansen et al., 1999; Sato et al., 2000); but as in other non-mammalian vertebrates there is a true class I region, displaying tight linkage of presenting and processing genes (reviewed by (Kulski et al., 2002)). Furthermore, MHC class I has been shown to evolve differently in various teleost species with large expansions (i.e. high number of expressed class Ia loci) reported e.g. in cod (Miller et al., 2002) and to a lesser extent in stickleback (Sato et al., 1998), while other species present a more conservative number of genes: two classical loci in medaka (Tsukamoto et al., 2005) and only one in salmonids (Grimholt et al., 2002; Shum et al., 2001). This major highly polymorphic locus, as well as both medaka loci, belongs to the U lineage, but several U lineage loci with little or no polymorphism (non-classical) are also present (Lukacs et al., 2010; Tsukamoto et al., 2005). Besides the broad U lineage, encompassing most published class I sequences from fishes, other three MHC class I loci-lineages have been identified in teleosts, namely: Z, L and S lineages (Dijkstra et al., 2007; Lukacs et al., 2010; Stet et al., 1998). The Z lineage has representatives from fewer teleosts, the L lineage only from cyprinids and salmonids, and S lineage only from salmonids and catfish (Lukacs et al., 2010). Remarkably, the different domains of the MHC class I also evolve differently (locus-specific, variable and concerted/species-specific manners for α1, α2 and α3 domains, respectively), and across-species lineages for the individual domains have been reported (Nonaka et al., 2011). Recently, several immune-related genes have been identified in European sea bass by generating ESTs (Sarropoulou et al., 2009), but neither β2m nor MHC class I genes or transcripts have been characterized in detail. In this study, the β2-microglobulin and several classical MHC class I HC molecules from sea bass, an important species for the Mediterranean fish-farming industry, were isolated and characterized for the first time. The sequence of the Dila-β2m gene, as well as its organization and copy-number were determined. Several transcripts of the MHC class I and two partial genes were also identified and their copy-number evaluated. Comparative sequence and phylogenetic analyses were performed for both proteins and the structure of the complex was predicted by 3D homology modeling. Furthermore the expression profile of both molecules was determined by real-time PCR and the expression of β2m was also analyzed in the head kidney of stimulated fish.

2. Materials and methods

2.1. Fish

Sea bass, *Dicentrarchus labrax*, were kept in a recirculating, ozone-treated salt-water (25–30‰) system at 22 ± 1 °C and fed with commercial pellets twice a day. Organs used for cDNA cloning and genomic DNA extraction were removed from fish previously sacrificed with a lethal dose of 2-phenoxyethanol (Pancreac; >5 mL/10 L). For the expression experiments sea bass juveniles (approx. 150 g) were obtained from an aquaculture farm ("Nuova Azzurro", Civitavecchia, Italy), kept at 15 ± 1 °C, and the organs removed as described above.

2.2. cDNA cloning

2.2.1. β2m

cDNA previously obtained from head kidney mRNA of one stimulated fish was used. Degenerate primers were designed based on conserved regions of several fish β2m amino-acid sequences available in the GenBank database. The cDNA was amplified by PCR with primers β2mFW1/β2mRV3 (Table S1). Reaction conditions were: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 51 °C for 1 min, 72 °C for 30 s; and 72 °C for 5 min. The amplification product (~160 bp) was purified (QIAquick Gel Extraction Kit, Qiagen), cloned (pGEM®-T Easy Vector Systems, Promega) and sequenced (Eurofins).

The resulting partial nucleotide sequence was used to design three specific reverse primers (DLβ2mRV1, DLβ2mRV2, DLβ2mRV3; Table S1) in order to obtain the 5′ untranslated region (UTR) of the molecule. Total RNA previously isolated from the head kidney of one stimulated fish was used to synthesize first-strand cDNA with Superscript III™ (Invitrogen) and primer DLβ2mRV1. After cDNA purification (Sephadex G-50 span column), the 5′ RACE System for Rapid Amplification of cDNA Ends (Invitrogen) was used to prepare dCTP-tailed cDNA. Tailed cDNA (2 L) was first amplified with primers AAP/DLβ2mRV2 (Table S1). The reaction was performed in the above-described conditions, except for annealing at 54 °C. A single PCR product of ~300 bp was obtained and later purified, cloned and sequenced as described above.

To obtain the full-length sea bass β2m cDNA, specific primers were designed at the beginning of the 5′ UTR. DLβ2mFW1 and APv2 (Table S1) were used in a first amplification PCR (cDNA synthesized with APv2 using the total RNA previously used in 5′ RACE) with the following conditions: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 52 °C for 1 min, 72 °C for 1 min; and 72 °C for 5 min. The ~1200 bp PCR product was purified, cloned and sequenced (3 independent clones).

2.2.2. Heavy chain

Degenerate primers were designed based on conserved regions of several fish HC amino-acid sequences available in the GenBank database. The cDNA previously used for obtaining β2m was amplified with primers MHCIFW1/MHCIRV1 (Table S1). Cycling conditions were: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 51 °C for 1 min, 72 °C for 30 s; and 72 °C for 5 min. The resulting product (~560 bp) was purified, cloned, and plasmid DNA extracted and sequenced as described above.

Using the partial nucleotide sequence obtained, specific reverse primers (DLMHCIRV1, DLMHCIRV2; Table S1) were designed in order to obtain the 5′ UTR. Total RNA previously isolated from the head kidney of one stimulated fish was now used to synthesize first-strand cDNA with Superscript III™ (Invitrogen) and primer DLMHCIRV1. After purification and dCTP tailing (as described above), the cDNA (2 L) was first amplified with primers AAP/DLMHCIRV1 (Table S1). The reaction was performed as previously described except for annealing at 52 °C, and extension for 90 s. A second PCR amplification was performed using DLMHCIRV2/AUAP primers adjusting annealing to 55 °C. Two PCR products (∼400 bp) were obtained and later purified, cloned and sequenced as described above.

To amplify full-length sea bass MHCI cDNA, specific primers were designed at the beginning of the 5′ UTR. DLMHCIFW3 (Table S1) and APv2 were used in a 50 L first amplification PCR (cDNA synthesized with APv2 using the total RNA previously obtained in the laboratory from the head kidney of one stimulated fish) in the following conditions: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 59 °C for 1 min, 72 °C for 2 min; and 72 °C for 5 min. A semi-nested PCR was performed with primers DLMHCIFW4/APv2 (Table S1), in the same conditions except for annealing at 58 °C. The ∼2500 bp PCR product was purified, cloned and sequenced (2 independent clones). Additional full-length cDNA clones of MHCI were obtained from the thymus of a different fish, using DLMHCIFW3/APv primers with annealing at 61 °C. The ∼2500 bp PCR product was purified, cloned and 12 independent clones sequenced. Finally, full transcripts from a third fish (HK cDNA) were also amplified using primers DLMHCIFW3/DLMHCIRV17 (Cyt-L1) or DLMHCIRV15 (Cyt-L2) (Table S1). The ∼2500 bp PCR products were purified, cloned and six independent clones from each were sequenced.

2.3. Genomic DNA cloning

Genomic DNA (gDNA) was isolated from erythrocytes from two different fish, as described by Stet et al. (1993). To obtain the full β2m gene, gDNA was amplified by PCR using primers DLβ2mFW1/DLβ2mRV4 (Table S1) in the following conditions: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 46 °C for 1 min, 72 °C for 2 min; and 72 °C for 5 min. The PCR product was purified, cloned and 3 independent clones were sequenced.

To obtain Dila-UA partial genes, gDNA (Fish 3) was amplified with primers DLMHCIFW18/DLMHCIRV22 (Cyt-L1) or DLMHCIFW8/DLMHCIRV14 (Cyt-L2) (Table S1). The PCR products of ∼2.3 and ∼3.9 kb, respectively were purified and directly sequenced.

2.4. Sequence analysis

Full nucleotide and protein sequences from sea bass β2m and HC were compared to several β2m and HC sequences currently available in the GenBank database [http://www.ncbi.nlm.nih.gov/genbank/]. The multiple sequence alignments were made using CLUSTAL W (Higgins, 1994) and formatted with Bioedit (Hall, 1998). Putative leader sequences were identified with SignalP 3.0 (Bendtsen et al., 2004). Boundaries of the Ig-like domain as well as of the Ig/MHC protein were predicted with InterPro Scan (Zdobnov and Apweiler, 2001). The transmembrane region was predicted with the Swiss Model domain annotation tool (Arnold et al., 2006). Molecular weights of the derived polypeptides were calculated with Expasy compute pI/MW tool [http://www.expasy.org/tools/pi_tool.html]. Putative N-glycosylation sites were predicted with the Expasy post-translational modification tool [http://www.cbs.dtu.dk/services/NetNGlyc/]. Neighbor-joining phylogenetic trees were constructed either with MEGA v5.05 (Tamura et al., 2011), using p-distance parameter and pairwise deletion of gaps or with ClustalX (Larkin et al., 2007), using all sites for separate domain analyses. All trees were tested for reliability using 1000 bootstrap replications. Percentages of similarity and identity were calculated with MatGAT (Campanella et al., 2003) using default parameters. Finally, ligand-type specificity of all sea bass MHC class I heavy chain molecules was predicted with MHCLIG [http://imed.med.ucm.es/MHCLIG/].

2.5. Southern blotting

Genomic DNA (∼10 g) was digested (ON at 37 °C) with different restriction enzymes (BamHI, EcoRI and HindIII for β2m, and EcoRI, NcoI and NdeI for MHCI HC all zero cutters within the probe), separated by electrophoresis in a 0.8% agarose gel and subjected to Southern blotting (Ausubel et al., 1999). Segments of the sea bass β2m (exon II) and of MHCI HC (exon IV) cDNAs, were amplified with primers DLβ2mFW4/DLβ2mRV7 and DLMHCIFW6/DLMHCIRV8 (Table S1), respectively, with conditions: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 54 °C/58 °C for 1 min (β2m/HC), 72 °C for 20 s; and 72 °C for 5 min. Both PCR products (273 bp for β2m and 246 bp for HC) were purified as described before, labeled and used as probes.

Preparation of the labeled probes, hybridization and post-hybridization stringency washes were performed accordingly to Gene Images™ AlhPhos Direct™ Labelling and Detection System (Amersham Biosciences). For signal generation and detection the Chemiluminescent Signal Generation and Detection with CDP-*Star*™ protocol from the same kit was followed.

2.6. 3D homology modeling

Suitable structural templates for Dila-β2m (grass carp β2m [PDB: 3gbl]) and Dila-UA (chicken MHC class I HC [PDB: 3bev]), were identified by a BLAST search as implemented in the SWISS-MODEL Protein Modelling Server (Arnold et al., 2006) and the automatic sequence alignments thus obtained were used for homology modeling with the same server. The calculated global energies were of −4362.397 kJ/mol (Dila-β2m) and of −9596.217 and −11186.899 kJ/mol (Dila-UA∗06 and ∗12, respectively). The templates used in model generation display identities of approximately 62% (Dila-β2m) and 36–38% (Dila-UA∗06 and ∗12, respectively). The secondary structure elements were identified with PROMOTIF (Hutchinson and Thornton, 1996). A dimeric model was generated by superposing the β2m and HC (Dila-UA∗06) homology models with the corresponding chains of the experimental chicken tertiary complex (Koch et al., 2007), and submitted to geometry regularization with PHENIX (Adams et al., 2010), with a resulting QMEAN score of 0.536 (Benkert et al., 2011). Model quality was assessed with WHATCHECK (Hooft et al., 1996) and PROCHECK (Laskowski et al., 1993). Pictures of the homology models were prepared with PyMOL [http://pymol.org/].

2.7. Quantitative PCR analyses of β2m and MHC class I

2.7.1. Basal expression analyses

To study the MHC class I andβ2m basal expression, six sea bass juveniles were sampled and leucocytes from different tissues (muscle, thymus, head kidney (HK), liver, spleen, gills, brain, peripheral blood leucocytes (PBL), gut) obtained as described previously (Scapigliati et al., 2001). Total RNA was isolated from each tissue separately with Tripure (Roche), resuspended in DEPC-treated water and used for real-time quantitative PCR. For reverse transcription, the BioScript RNase H minus (Bioline) enzyme was used as previously described (Buonocore et al., 2007). The expression levels of the selected transcripts were determined with a

Mx3000P™ real time PCR system (Stratagene) equipped with version 2.02 software and using Brilliant SYBR Green Q-PCR Master Mix (Stratagene) following the manufacturer's instructions, with ROX as internal passive reference dye. Specific PCR primers were designed for the amplification of about 200 bp products from MHC class I (in the conserved region of all known sea bass MHC class I sequences), β2m, and from 18 S ribosomal RNA (Table S1), the later used as a house-keeping gene. Ten ng of cDNA template were used in each PCR reaction, with conditions: 95 °C for 10 min, followed by 35 cycles of 95 °C for 45 s, 52 °C for 45 s and 72 °C for 45 s. Triplicate reactions were performed for each template cDNA (replaced with water in all blank control reactions). The analysis was carried out using the endpoints method option, with collection of the fluorescence data at the end of each extension stage of amplification. A relative quantitation has been performed, comparing the levels of the target transcripts (MHC class I and β2m) to a reference transcript (calibrator, the tissue with the lowest expression, in this case muscle). A normaliser target (18S ribosomal RNA) is included to correct for differences in total cDNA input between samples. The results are expressed as the mean ± SD of the results obtained from the six individuals. The real-time PCR products from the different tissues were successively examined by agarose gel electrophoresis to investigate their specificity and size and sequenced to verify the correct amplification.

2.7.2. β2m expression analysis after in vitro stimulation

In vitro expression studies were performed using head kidney leukocytes (Scapigliati et al., 2001) isolated from six juvenile sea bass (150 g weight) cultured in L-15 medium (Gibco) containing 10% FCS, adjusted to 1×10^5 cells/ml and incubated at 18 °C for 6 h and 24 h with 50 g/ml of poly I:C (Sigma). The control cell sample was stimulated with L-15 alone. Total RNA was isolated with Tripure (Roche), resuspended in DEPC-treated water and used for real-time quantitative PCR. The primers and real time PCR conditions were the same as described in the previous section, except that the calibrator for this experiment was the time 0 control. The results were expressed as the mean ± SD of the results

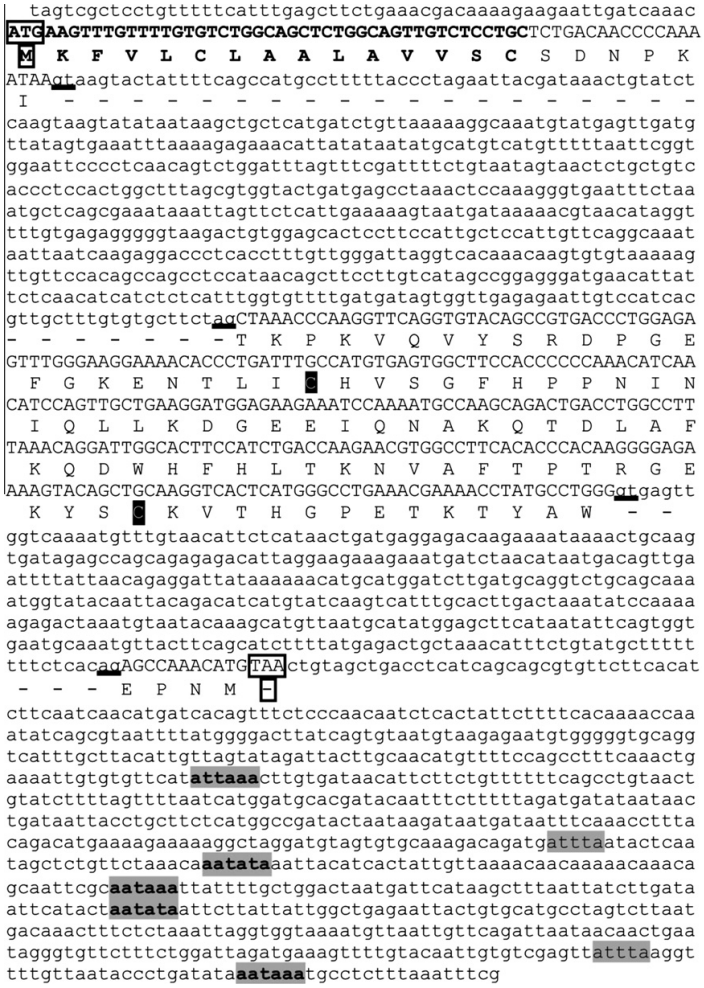


Fig. 1. Nucleotide and deduced amino acid sequences of sea bass β2m gene. The ORF and the predicted protein are shown in uppercase letters. The translation start and stop codons and the initial methionine are boxed. The predicted signal peptide is highlighted in bold, and cysteines are in inverted type. Within the 3′ UTR, the putative instability motifs are shaded in grey. The poly-adenylation signal sequences are in bold and shaded in grey. The intron splicing consensus sequences (gt/ag) are underlined.

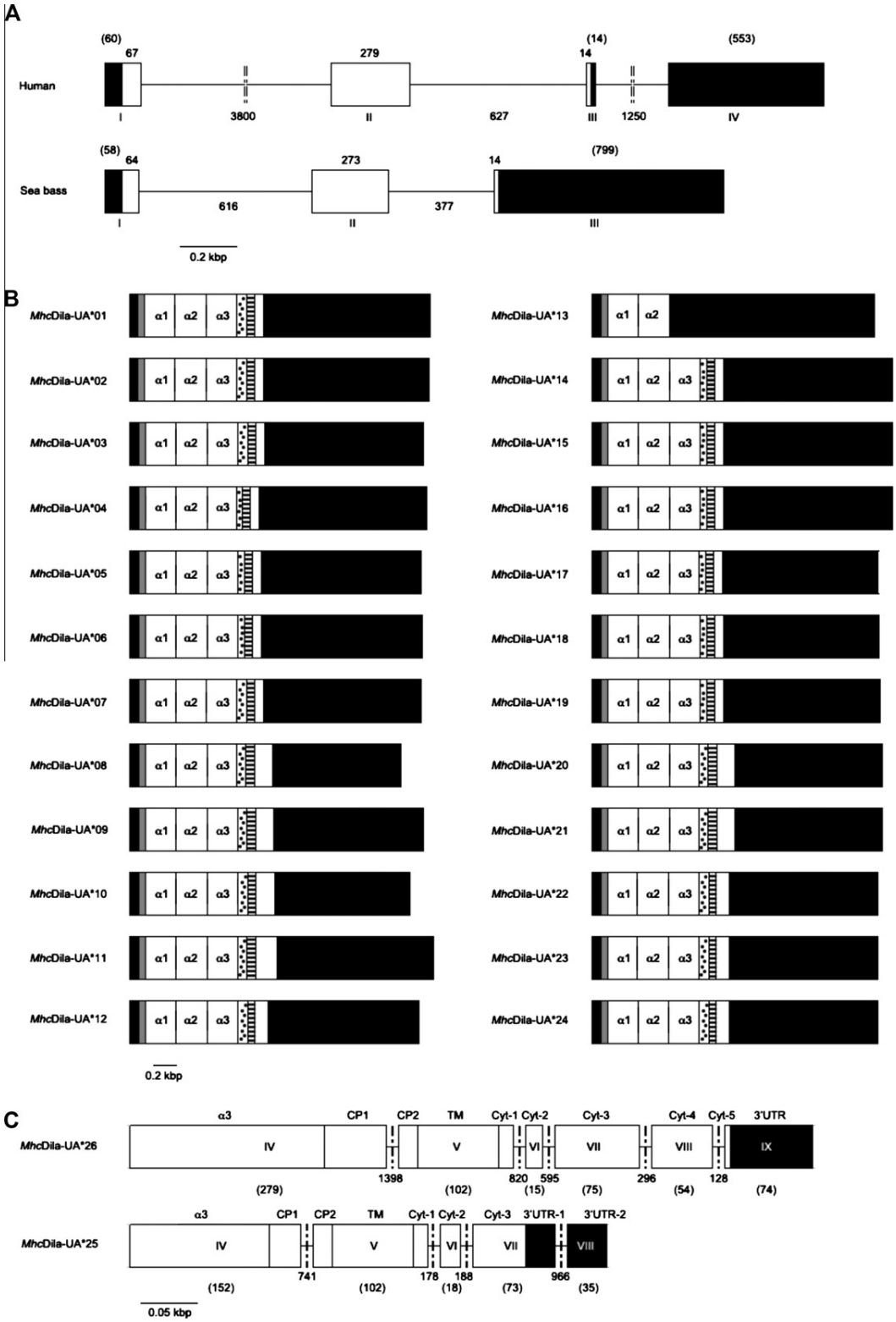


Fig. 2. (A) Schematic representation of sea bass and human β2m gene structures. Boxes represent exons and horizontal lines introns. White and black boxes differentiate coding and untranslated regions, respectively. Values above boxes and below the lines represent the number of nucleotides (size) of exons and introns, respectively. Exon numbers are indicated below boxes by Roman numerals. (B) Schematic representation of the structures of sea bass MHC class I heavy chain full transcripts. Black bars represent 5′ and 3′ UTRs, while the different exons/domains within the MHC class I heavy chain coding region are represented by patterned/labeled bars: grey – signal peptide; α1–α1 domain; α2–α2 domain; α3–α3 domain; black circles – connective peptide; horizontal lines – transmembrane domain; white – cytoplasmic tail. The name of each transcript is indicated on the left. (C) Schematic representation of two Dila-UA partial gene structures. Boxes represent exons and horizontal lines introns. White and black boxes differentiate coding and untranslated regions, respectively. Values below boxes and lines represent the number of nucleotides of exons and introns, respectively. Putative exon numbers are indicated inside boxes by Roman numerals. Boxes are labeled according to the distinct domains they codify: α3–α3 domain; CP, connective peptide; TM, transmembrane domain; Cyt, cytoplasmic tail; 3′ UTR–3′ UTR. Please note that for both genes the α3 and the 3′ UTR regions are incomplete.

obtained from six fish and the differences were considered significant if $p < 0.05$ using the two-way ANOVA analysis followed by Bonferroni's post test.

3. Results and discussion

3.1. Transcripts and gene sequences analysis

3.1.1. β2-Microglobulin

The full-length sea bass β2m (*Dila-β2m*) cDNA was obtained by RT-PCR using mRNA extracted from the head kidney of one fish. BLAST analysis confirmed highest similarity with fish β2m. The here described *Dila-β2m* sequence [GenBank: HQ290128] has a total size of 1208 bp, with a 58 bp 5′ UTR, and a 799 bp 3′ UTR (excluding the poly(A)-tail). Within the 3′ UTR there are five potential polyadenylation signals, the last of which 17 nucleotides upstream from the actual poly(A)-tail, as well as two ATTTA instability motifs, believed to be responsible for rapid mRNA degradation (Shaw and Kamen, 1986). These features are also observed in the here-reported *Dila-β2m* gene [GenBank: HQ290127] (Fig. 1), which displays a three-exon structure.

In the *Dila-β2m* gene all intron donor/acceptor sites match the consensus motifs (gt/ag), and the codons in the exon boundaries are split by introns between the first and second nucleotides (Fig. 1). Exon I encodes the 5′ UTR, the putative leader peptide, the first six amino acids of the mature protein and the first nucleotide of codon 7; exon II covers the remaining two nucleotides of codon 7 and the IgSf C domain to the first base of codon 113; finally, exon III encodes the remaining two bases of codon 113, the three-last amino acids of the protein and the 3′ UTR (Fig. 1). Both features of the exon/intron pattern from the sea bass β2m gene, i.e. phase-one introns and correspondence of exons to functional domains of the protein, are typical in members of the Ig superfamily (Williams and Barclay, 1988). Morevoer, the *Dila-β2m* 3-exon gene organization (Fig. 2A) is shared by several other fish species, namely sturgeon (Lundqvist et al., 1999), walleye (Christie et al., 2007), grass carp (Hao et al., 2006), catfish (Criscitiello et al., 1998), tongue sole (Xu et al., 2010), and sandbar shark (Chen et al., 2010a). These, present a genomic structure different from that of human β2m, which has four exons (Gussow et al., 1987). In human β2m, besides the last four amino acids of the protein and the stop codon, exon III includes only the first 14 nucleotides of the 3′ UTR, being the remaining portion specified by exon IV. The 4-exon structure of human β2m gene is also found in mouse (Parnes and Seidman, 1982), chicken (Riegert et al., 1996) and zebrafish (Ono et al., 1993). Japanese flounder β2m gene has been reported to contain a single exon (Choi et al., 2006).As in other fish β2m, introns are smaller than in those of the human gene (Fig. 2A).

3.1.2. MHC class I heavy chain

Two complete cDNA sequences from one individual, 11 from a second fish and 11 others from a third one were identified and designated *MhcDila-UA∗01-MhcDila-UA∗24* [GenBank: HQ290103–HQ290115; and JX171686–JX171696], after sequence analysis had confirmed highest similarity to fish MHC class I molecules, skipping only the middle letter locus designation according to Klein and coworkers nomenclature (Klein et al., 1990). *MhcDila-UA∗13* contains a 5′ UTR and an ORF comprising only the signal peptide and the peptide-binding region (PBR) –α1 and α2 – domain (Fig. 2B) due to the transcription of an intron that prematurely introduces a stop codon. Therefore it most likely represents a splice variant, as in absence of the intronic sequence, the coding region would continue through the α3 to the transmembrane domain prematurely ending before the full cytoplasmic tail (possibly due to the presence of several adenosines to which the oligod (T) paired, originating a truncated transcript), or an expressed pseudogene. All the remaining sequences (Fig. 2B) are composed of 5′ and 3′ UTRs, as well as of signal peptide, α1, α2, and α3 domains, and of connective peptide/transmembrane/cytoplasmic tail (CP/TM/CYT) regions, and were used for further analysis.

The here reported sea bass class I full transcripts have total sizes of 2404–2683 bp, with 5′ UTRs of 66–83 bp (differences resulting from different primer usage), ORFs of 1068–1215 bp and 3′ UTRs of 1141–1500 bp excluding polyadenylation tails (Table 1). Expected variability is denoted at the nucleotide level on the PBR domains (Figs. S1 and S2). However, striking differences at nucleotide composition and length of the CP/TM/CYT and 3′ UTR regions (Figs. S1 and S2 and 2B) segregate them into two main groups. One is more homogeneous being characterized by generally short CPs (20 residues; three exceptions), constant TM regions (23 residues), short CYT tails (25 residues), and longer 3′ UTRs with 11–15 potential poly(A) signals, 6–7 ATTTA instability motifs and two stretches of CA dinucleotide repeats [$(CA)_n$] of variable lengths (polymorphism; Table 1 and Fig. S1). The significance of these putative microsatellites remains to be investigated, although the presence of such a feature has been previously reported in Atlantic salmon (Grimholt et al., 1993).

The other group of transcripts is represented by long CPs (29 residues), heterogeneous TM regions (20–25 residues) and CYT tails (27–62 residues), and 3′ UTRs with 4–9 poly(A) signals, generally 9 ATTTA motifs and no microsatellites (Table 1and Fig. S2). Variations at the TM domains are due to predicted boundaries and not to indels – insertions/deletions – along the domain, while those at the CYT tails starkly differ in size.

Nevertheless, both groups of European sea bass Dila-UA transcripts have a predicted leader peptide of 19 amino acids and, with few exceptions, α1, α2, and α3 domains of 89, 92, and 90 amino acids, respectively (Table 1). Hence, it can be inferred that differences in length/size among European sea bass MHC class I UA sequences result essentially from distinct CP/TM/CYT regions (Fig. 2B and Table 1).

Partial genes from each of these groups were isolated in order to elucidate if some of the observed transcripts could be splice variants. These sequences, corresponding to regions from the α3 to the 3′ UTR, were named *MhcDila-UA∗25* and *∗26* [GenBank: JX171697 and JX171698]. Likely, transcript Dila-UA∗16 originates on gene *Dila-UA∗25*, while Dila-UA∗21 on gene *Dila-UA∗26*, since they were obtained from the same fish and there is 100% identity on coding regions. The remaining transcripts may be allelic forms or originate on different genes. In both *Dila-UA* partial genes all intron donor/acceptor sites flanking the exons match the consensus motifs (gt/ag), and the codons in the exon boundaries are split by introns between the first and second nucleotides; however they differ not only in size (2453 and 3835 bp for short and long CYT tails, respectively), but also in gene organization (number and length of exons/introns) (Fig. 2C). Dila-UA∗25 has five exons while Dila-UA∗26 has six, with the CPs and the CYT tails encoded by more than one exon (Fig. 2C). Although the cytoplasmic tails are differently organized in three or five exons, no alternative splicing sites could be identified in the current transcripts, even within the more heterogeneous long CYT tails group.

3.2. Southern blot

Southern blot results suggest that a single gene encodes Dila-β2m in sea bass (Fig. 3A). Similar results have been reported in other organisms including mammals (Parnes and Seidman, 1982), birds (Riegert et al., 1996), amphibians (Stewart et al., 2005), cartilaginous (Chen et al., 2010a) and bony fish (Choi et al., 2006; Christie et al., 2007; Criscitiello et al., 1998; Dixon et al., 1993; Ono et al., 1993). In some species, a higher copy-number of the β2m gene is

Table 1
Sea bass MHC class I heavy chain clones characteristics.

Fish	MHCI molecule	Transcripts							Proteins										
		length (nucleotides)				Microsatellite [CA]*n*	Poly(A) Signals	mRNA instability attta	Length (amino acids)										
		Full transcript	5′ UTR	ORF	3′ UTR				Precursor	MW (kDa)	Signal peptide	Mature protein	MW (kDa)	α1 domain	α2 domain	α3 domain	CP	TM (3)	Cyt tail
1	MhcDila-UA*01	2652	66	1104	1482	⌐⌐(1)	12	7	367	41.7	19	348	39.8	89	92	90	29	23	25
	MhcDila-UA*02	2643	66	1104	1473	⌐⌐	12	7	367	41.7	19	348	39.8	89	92	90	29	23	25
2	MhcDila-UA*03	2665	81	1104	1408	⌐⌐	12	7	367	41.7	19	348	39.8	89	92	90	29	23	25
	MhcDila-UA*04	2640	81	1068	1491	⌐⌐	13	7	355	40.3	19	336	38.4	89	92	88	18	24	25
	MhcDila-UA*05	2578	81	1077	1420	⌐⌐	15	6	358	40.4	19	339	38.4	89	92	90	20	23	25
	MhcDila-UA*06	2586	81	1077	1428	⌐⌐	13	7	358	40.5	19	339	38.5	89	92	90	20	23	25
	MhcDila-UA*07	2585	78	1104	1403	× (2)	6	10	367	41.7	19	348	39.7	89	90	90	29	23	27
	MhcDila-UA*08	2404	81	1182	1141	×	6	7	393	44.1	19	374	42.2	89	92	88	27	25	53
	MhcDila-UA*09	2603	83	1191	1329	×	6	9	396	44.5	19	377	42.6	89	92	90	27	25	54
	MhcDila-UA*10	2479	81	1200	1198	×	4	5	399	44.7	19	380	42.8	89	92	90	29	22	58
	MhcDila-UA*11	2683	81	1215	1387	×	9	9	404	45.8	19	385	43.9	89	92	90	27	25	62
	MhcDila-UA*12	2561	81	1137	1343	×	7	9	378	43.1	19	359	41.2	89	92	89	29	20	40
	MhcDila-UA*13	2489	78	618	1811	×	6	3	205	–	19	186	–	89	92	–	–	–	–
3	MhcDila-UA*14	2657	81	1077	1499	⌐⌐	12	7	358	40.5	19	339	38.6	89	92	90	20	23	25
	MhcDila-UA*15	2658	81	1077	1500	⌐⌐	12	7	358	40.4	19	339	38.5	89	92	90	20	23	25
	MhcDila-UA*16	2654	81	1077	1496	⌐⌐	12	7	358	40.5	19	339	38.6	89	92	90	20	23	25
	MhcDila-UA*17	2541	81	1071	1389	⌐⌐	12	6	356	40.2	19	337	38.3	89	92	88	20	23	25
	MhcDila-UA*18	2542	81	1077	1384	⌐⌐	13	6	358	40.7	19	339	38.7	89	92	90	20	23	25
	MhcDila-UA*19	2550	81	1077	1392	⌐⌐	11	6	358	40.7	19	339	38.7	89	92	90	20	23	25
	MhcDila-UA*20	2568	81	1182	1305	×	4	9	393	44	19	374	42.1	89	92	88	27	25	53
	MhcDila-UA*21	2568	81	1182	1305	×	4	9	393	44.1	19	374	42.1	89	92	88	27	25	53
	MhcDila-UA*22	2531	81	1134	1316	×	5	9	377	42.6	19	358	40.6	87	92	90	29	20	40
	MhcDila-UA*23	2531	81	1134	1316	×	5	9	377	42.7	19	358	40.7	87	92	90	29	20	40
	MhcDila-UA*24	2528	81	1131	1316	×	5	9	376	42.8	19	357	40.9	89	90	89	29	20	40

(1) Each (⌐) symbol represents a microsatellite. (2) The symbol (×) denotes absence of microsatellite. (3) TM predictions of Swiss Model domain annotation tool.

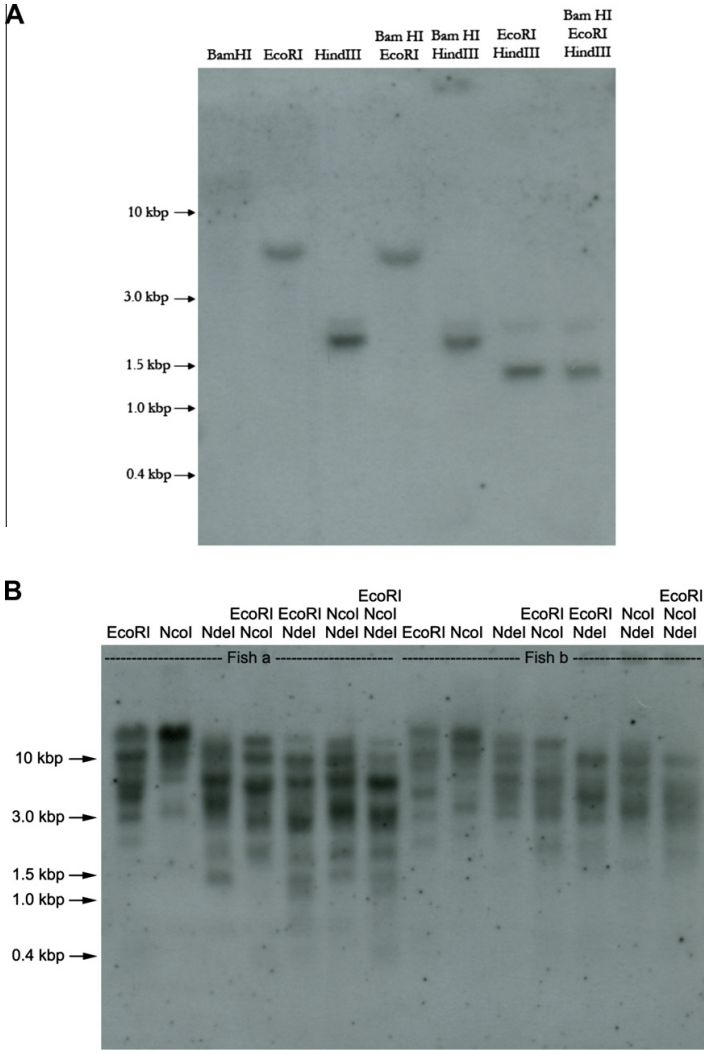


Fig. 3. (A) Sea bass β2m is a single copy gene. Southern blot analysis of genomic DNA from one animal digested with three restriction enzymes (isolated and combined). The blot was later probed with a region of β2m that included part of exon II. All enzymes are zero cutters within the probe: one band can be seen for each of these individual digestions and for each set. Sizes of the MW marker are indicated on the left. (B) Sea bass MHC class I heavy chain is a multi-copy gene. Caption as for β2m, except that results from two different animals are shown and the probe includes exon IV. Multiple bands can be seen for each of the individual digestions and for each set.

explained by the polyploidy of genomes: sturgeon (Lundqvist et al., 1999), common carp (Dixon et al., 1993) and African clawed frog (Stewart et al., 2005) are tetraploid species, displaying two copies of the β2m gene, while barbs being hexaploid have three copies of the gene (no diploidization) (Kruiswijk et al., 2004). In cod, however, which is a diploid species, two fragments were also identified but were suggested to be alleles of a single gene (Persson et al., 1999). Limited polymorphism of this molecule was long reported in mice (Goding and Walker, 1980). Nevertheless, in rainbow trout, sequence diversity in β2m cDNA clones from a single fish has been reported (Shum et al., 1996) and is now known that β2m is encoded by three genes segregating as tetraploid loci (Magor et al., 2004).

For *Dila-UA*, using an α3 domain-specific probe, multiple bands (5–8) were observed upon southern blot analysis of two sea bass genomes (Fig. 3B). This result determines MHC class I molecules are polygenic, although the actual number of class I loci in European sea bass is still not clearly defined; it is also suggestive of haplotypic variation, since the pattern is not identical in the analysed individuals. Polygenic class I molecules have been reported in mammals, birds, amphibians, cartilaginous and bony fish (reviewed in (Kelley et al., 2005)) and haplotype variation is also common among teleost fish (Figueroa et al., 2001; Michalova et al., 2000; Murray et al., 2000). Co-dominance is another important feature of MHC class I alleles, but there are species where a single gene is dominantly expressed (e.g. in salmonids (Aoyagi et al., 2002; Miller et al., 2006), African clawed frog (Flajnik et al., 1999) and chicken (Kaufman et al., 1999)). Although it is difficult to determine from our data whether the different cDNA clones are alleles or the products of different genes, one can say that at

least six different MHC class I genes must occur in sea bass since eleven different transcripts were obtained in each of two individuals (Table 1). This is in accordance to the 5–8 bands observed on southern blot, which suggests a total of 10–16 sequences may exist per fish. Therefore at least two genes may still be undiscovered.

3.3. Protein sequence comparison

3.3.1. β2-Microglobulin

Amino acid sequences of β2m from several vertebrate species (Table S2) were aligned with that of Dila-β2m using CLUSTALW (Fig. 4). In sea bass, a predicted 101-residue (11.6 kDa) mature protein would be generated by cleavage of the putative 15-amino acid signal peptide. Although the majority of the analyzed bony fish species have a 97-residue β2m (two amino acids shorter than in human, mouse and amphibians), in other members of the supraorder Percomorpha it is 100-residue long. The cartilaginous fish display smaller mature proteins of 94 or 95 amino acids. Sea bass β2m, as all other fish species, except for sturgeon, has a deletion of two residues in the region corresponding to the F–G loop. The IgC1 motif and the Ig/MHC protein signature (sb residues 82–88; [FY]XCX[VA]XH) (Fig. 4), and the canonical pair of cysteines (hC^{25}–C^{80}, sbC^{29}–C^{84}) covalently linking the IgSf domain through strands B and F in human β2m (Bjorkman et al., 1987b), are conserved across species. As in other β2m proteins, no putative glycosylation sites are present on the sea bass sequence, although they were predicted for a few fish species (Criscitiello et al., 1998; Hao et al., 2006; Yu et al., 2009). Overall, conservation can be observed between sea bass and human β2m sequences (Fig. 4), especially in secondary structure elements (Saper et al., 1991) (least conserved regions include β-strands E and G and loop F–G), and among residues involved in interactions with MHC class I α1, α2 and α3 re-

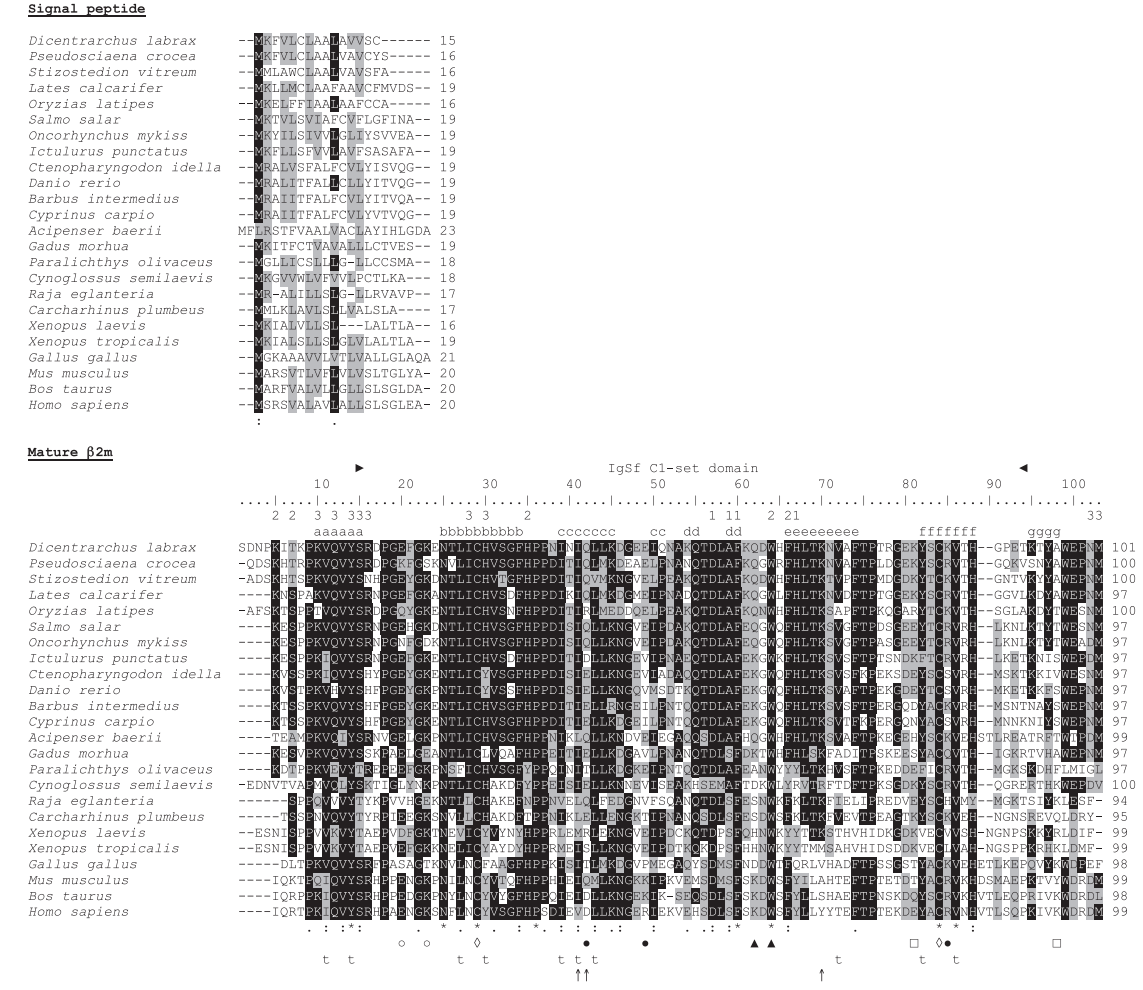


Fig. 4. Alignment of β2m amino acid sequences. Amino acid sequences were retrieved from GenBank (Table S2) and aligned with CLUSTALW. Dashes indicate gaps that maximize the alignment; Identical residues, conserved and semi-conserved substitutions are denoted under the alignment with, (∗), (:) and (.), respectively. Identity and similarity are displayed by black and grey shading, respectively. Predicted IgSf C domain is indicated above the alignment. Canonical cysteines involved in an intra-chain disulfide bond (B–F) are marked with (◇). Four pairs of salt bridges are indicated with (○, □ and ●). Residues of human β2m that contact CD8 (Gao et al., 1997) are indicated with (▲). Topohydrophobic residues are indicated with (t). Residues involved in two new hydrogen bonds in grass carp are indicated with (). Letters (a–g) above the alignment denote β-strands (Saper et al., 1991). The numbers 1, 2, and 3 above the alignment indicate the contacts of human β2m with heavy chain α1, α2, and α3, respectively (Saper et al., 1991).

gions (Tysoe-Calnon et al., 1991): globally 11/18 conserved and 4/18 conservatively substituted residues (α1: 2/4 conserved and 2/4 conservatively substituted; α2: 3/5 conserved; α3: 6/9 conserved and 2/9 conservatively substituted). Of note hD^{53}/sbD^{57}, involved in important hydrogen bonds to the HC α1 domain, is conserved across species, except for tongue sole where it is conservatively replaced by glutamate. Also residues hH^{31}, hW^{60}, and hF^{62} (sbH^{35}, sbW^{64} and sbF^{66}), known to form hydrogen bonds to the α2 domain, are totally preserved in sea bass and highly conserved across species. Finally, human Q^{8}, Y^{10}, R^{12} and N^{24} (sea bass Q^{12}, Y^{14}, R^{16} and I^{28}) form hydrogen bonds while Y^{26} (sea bass H^{30}) establishes hydrophobic interactions to the HC α3 domain. The non-conservative replacement of N^{24} by I^{28} in sea bass (note that isoleucine is present in bony fish and *Xenopus*; Fig. 4) leads to the loss of this interaction since the asparagine side-chain is involved. Conservation of the remaining residues as well as of their HC partners means those interactions are plausible in the sea bass MHC class I complex, as happens in most characterized species. The contacts with CD8α1, when cytotoxic T lymphocytes recognize target cells, are made through residues hK^{58} and hW^{60} (Gao et al., 1997), which in sea bass are conservatively substituted (Q^{62}) and conserved (W^{64}), respectively. Four salt bridges have been described in HLA-bound human β2m (Saper et al., 1991): E^{16}/K^{19}, D^{38}/R^{45}, D^{38}/R^{81}, and E^{77}/K^{94}. Only the first one is preserved in sea bass (E^{20}/K^{23}) and highly conserved across species. Two new hydrogen bonds involving K^{66} have been described in grass carp β2m connecting strands C–E (Chen et al., 2010b). This lysine is characteristic of fish β2m (replaced by arginine in tongue sole), being absent in *Xenopus tropicalis* (M), chicken (V), mouse (A), bovine (S) and human (Y) (Fig. 4). The environment of K^{66} in grass carp (K^{70} in sea bass) is maintained, allowing the establishment of the hydrogen bonds described for fish β2m, namely I^{37}/K^{66} (sb I^{41}/K^{70}) and E^{38}/K^{66} (sb Q^{42}/K^{70}). The typical hydrophobic core described for the Ig fold, with 10 topohydrophobic positions indicated as key markers of the C1 set (Halaby et al., 1999) is mostly conserved in sea bass. Finally, over the complete β2m sequence, eight residues are totally conserved and ten other show >90% conservation across species (Fig. 4).

3.3.2. MHC class I heavy chain

MHC class I amino acid sequences from several vertebrate species (Table S2) were aligned with some Dila-UA sequences using CLUSTALW (Fig. 5). Cleavage of the putative 19-amino acid signal peptides, and in accordance to what was previously discussed, would yield 336–385-residue mature proteins. Among all analyzed species the sea bass clone Dila-UA∗11 originates the longest protein and the African barb the shortest. The sea bass transcripts differed by 48/89 residues in the α1 domain, 55/92 in the α2 domain and 22/90 in the α3 domain (Fig. S3). The α1 and α2 domains of Dila-UA display high diversity. As mentioned above, unexpected and striking variability is also observed within the CP/TM/CYT (Fig. 5 and Fig. S3) yielding two groups of MHC class I molecules (short and long CYT tails) with distinct molecular weights (Table 1).

The insertion of a unique fragment – WGKTGIR[ST][EN] – in the CP of all sea bass transcripts with long CYTs and of three with short CYTs (Dila-UA∗01-∗03) makes them longer (Fig. 5 and Fig. S3). A similar insertion is also observed in *Sparus aurata* and *Takifugu rubripes* sequences (Fig. 5). No TM is predicted in the *Gasterosteus aculeatus* sequence, which possibly corresponds to a secreted form of the protein. In all other species here analyzed there is a TM region, including all sea bass sequences (Fig. 5 and Fig. S3). Six Dila-UA molecules (∗07, ∗11, ∗12, ∗22, ∗23 and ∗24) display at their long cytoplasmic tail a tyrosine equivalent to that (Y^{320}) reported to be involved in endolysosomal traffic of mammalian MHC class I complexes during cross-presentation of exogenous peptides in dendritic cells (Lizee et al., 2003). However, although conserved in several distantly related species, this residue is absent in most here-analyzed Actinopterygii HC molecules. Furthermore, the presence of leucine–isoleucine motifs within the long cytoplasmic tail molecules (Dila-UA∗08 to ∗12 and Dila-UA∗20 to ∗24) might also target them to endosomes, which would resemble MHC class II transport (Pieters et al., 1993). All Dila-UA characteristic long CYTs contain one or two cysteines (Fig. 5 and Fig. S3). Existence of cysteines on the cytoplasmic tails of MHC class I molecules (also observed in 10 of the analyzed species) has been reported to affect the expression level and conformation of the proteins at the cell surface in humans, and also to influence the binding/recognition of killer cell inhibitory receptors (Barel et al., 2003). For all this, it is conceivable that these two groups of European sea bass MHC class I molecules with distinct specificities at the CYT region may have different functions.

In spite of these segregating differences, both groups of sea bass protein sequences showed the typical features of classical MHC class I α chains when aligned with those of other species (Fig. 5 and Fig. S3). One pair of cysteines in the α2 domain (C^{100}/C^{164}) and another in the α3 domain (C^{200}/C^{260}), responsible for two intrachain disulfide bonds in the human molecule, could be identified. Exceptionally, some of the sea bass transcripts show one (C^{67}) or two cysteine residues (C^{35}/C^{67}) in the α1 domain (Dila-UA∗07, ∗10, ∗11 and ∗12; Fig. 5 and Fig. S3) as previously reported in medaka (Nonaka and Nonaka, 2010). An additional cysteine residue (in a position equivalent to C^{67}) has been reported in human HLA-B27, an allele associated with an autoimmune disorder (Antoniou et al., 2004).

A potential N-glycosylation site (N^{86}QT) was identified in a conserved position (Fig. 5 and Fig. S3). Site-directed mutagenesis of the MHC class I HC (mouse) determined that the glycosylated residue on the α1 domain is the binding site for calreticulin (CRT) (Turnquist et al., 2002), raising the possibility that a similar interaction occurs in sea bass HC.

Within the PBD, eight sites (YYRTKWYY) have been identified as important anchor residues for peptide binding (based on the mammalian motif YYYTKWYY), almost all of which are conserved in sea bass MHC class I molecules (Fig. 5 and Fig. S3): Y^{7} (I^{7} in Dila-UA∗04; Fin *Gadus morhua*; A in *Danio rerio* and *Channa argus*), Y^{59}, R^{84} (Y in mammals and iguana,), T^{142} (I in *G. aculeatus*,), K^{145} (R in *G. morhua*), W^{146} (L in Dila-UA∗05, ∗08, ∗17, ∗20-∗23, *Aulonocara hansbeanschi*, *C. argus*, *Homo sapiens* HLA-C), Y^{159} (D in *Epinephelus coioides*), and Y^{171} (F in *Protopterus aethiopicus*). Additional important residues for peptide binding (located at the bottom of the groove) (Bjorkman et al., 1987a) that are reasonably conserved include L^{5}, Y^{9}, F^{21} (Y in Dila-UA∗10-∗12) and G^{25} (S in Dila-UA∗04) (Hansen et al., 1996). Also F^{122} (I^{122} in Dila-UA∗08, ∗20 and ∗21; Y in mammals and iguana and also ancient Actinopterygii and *Heterodontus francisci*), L^{160} (Y in Dila-UA∗01-∗03) and W^{167}, part of the peptide-binding site (Saper et al., 1991), are well conserved in sea bass and across species.

Several structurally important residues of the α3 domain (Williams and Barclay, 1988), apart from the conserved cysteines, are totally conserved in the sea bass sequences: F^{205}, Y^{206}, P^{207}, W^{214}, G^{236}, Y^{258} and F^{262} (phenyalanine in all Percomorpha, although valine in most other species) are absolutely or highly conserved across species.

Important residues establishing contacts with β2m (Koch et al., 2007) are highly conserved across species including sea bass (F^{8}, T^{10}, V^{22}, Q^{31} (except ∗24), K^{45}, Q^{95}, Q^{114}, G^{116}, D^{118}, G^{119}, D^{121}, K^{192}, E^{229}, D^{235} and Q^{239}), while others are conservatively substituted in the later (V^{24}, E^{120}).

Additionally, a series of acidic residues thought to be critical for CD8 binding on cytotoxic T cells during antigen recognition by T cell receptors in mammals (Gao et al., 1997; Kern et al., 1998;

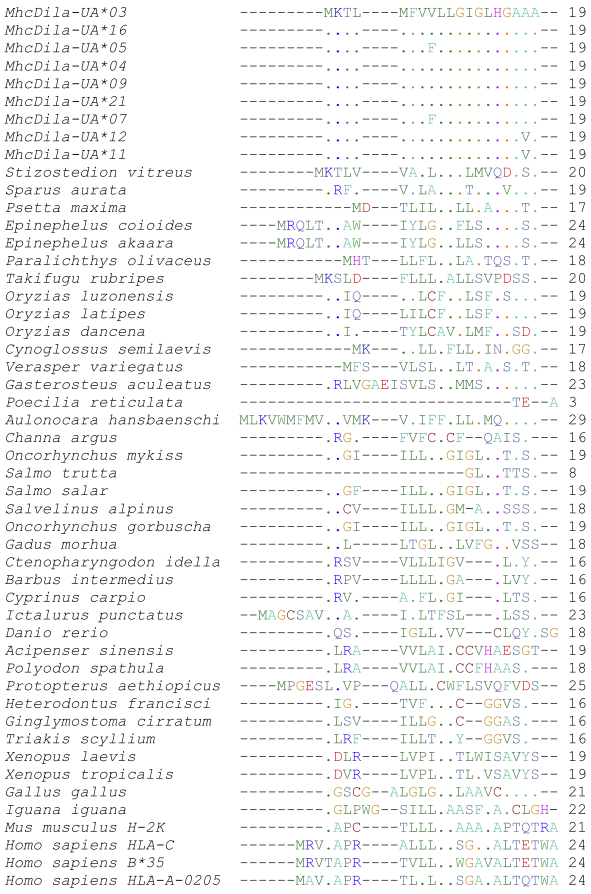


Fig. 5. Alignment of MHC class I heavy chain amino acid sequences. Amino acid sequences were retrieved from GenBank (Table S2) and aligned with CLUSTALW. Dashes indicate gaps that maximize the alignment, and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with, (∗), (:) and (.), respectively. Numbers right to the sequences denote their lengths; those within parenthesis indicate the size of individual domains. Canonical cysteines involved in an intra-chain disulfide bonds within α2 and α3 domains as well as other cysteines mentioned in the text are displayed in inverted type. Conserved residues located at the edges of the binding cleft are highlighted with red shading. Other important residues within the PBD are marked with (▼). Residues involved in salt bridges are denoted with green shading. Residues that contact β2m are highlighted with ($). Residues that contact CD8 (Gao et al., 1997; Kern et al., 1998; Wang et al., 2009) are indicated with yellow shading. A grey bar above the alignment denotes residues involved in tapasin contacts. For each sequence an open box delimits the TM domain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Wang et al., 2009) are also found in the α3 domains of the sea bass clones – E^{223} [DN]V[DE]HGE^{229} – although the identities are low between the various species (Fig. 5), suggestive of species-specific CD8 binding, as previously proposed (Kaufman et al., 1994). Other CD8-contacting residues include S^{209}, T^{211} and Q^{263} also in the α3 domain (Gao et al., 1997; Kern et al., 1998; Wang et al., 2009), conserved in the sea bass sequences but not across species; and Q^{114} and D^{121}(Gao et al., 1997; Koch et al., 2007) highly conserved in the α2 domains across species.

Residues believed to be involved in electrostatic interactions in the human molecule (Saper et al., 1991) are present in identical locations in the sea bass clones (Fig. 5 and Fig. S3). All these residues are invariant across species, except for D^{28} replaced by a glycine in *Poecilia reticulata* and an asparagine in *S. aurata* and *Salvelinus alpinus*. Another pair of residues in the α3 region involved in a salt bridge in human MHC seems to be present in most species (sb D^{217}/R^{257}) (Saper et al., 1991).

It has been demonstrated by co-precipitation experiments that polymorphisms at position 116 of HC α2 domain in both HLA-A and -B molecules affect binding to CRT, TAP and TPN (Turnquist et al., 2002). Although speculative, variability at the corresponding position is also observed in sea bass (sb^{115}: W/Y/D/F), raising the possibility of different affinities between distinct sea bass MHC class I alleles and the PLC members.

The regions of MHC class I HC that seem to interact with TPN reside in the α2 and α3 domains (Turnquist et al., 2002). These include a highly conserved solvent-exposed region (loop 128–136, sb L^{127}KTETFIAP) and the α3 region that also participates in CD8

α1 domain

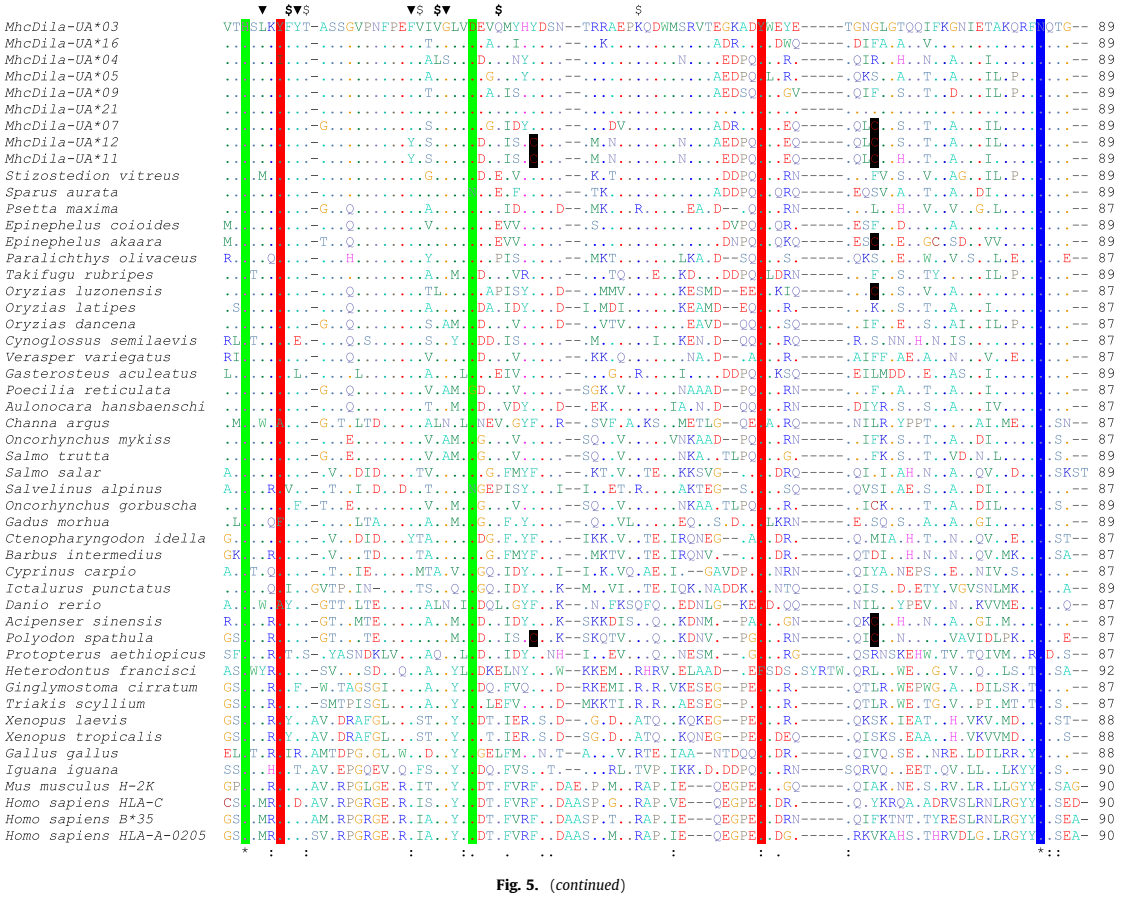


Fig. 5. (*continued*)

binding (sb 219–226) (Van Hateren et al., 2010). Namely S^{132} (α2 domain) and D^{227} and E^{229} (α3 domain) were found to affect association with TPN. Human S^{132} corresponds to the totally conserved T^{131} in sea bass. Regarding the α3 domain residues, human D^{227} is equivalent to D^{224} or N^{224} with similar frequencies. Human E^{229} corresponds to sea bass E^{226} or frequently D^{226}. The conservative nature of the replacements in sea bass, when at all, suggests the two interactions are maintained, a hypothesis that needs further experimentation.

Overall, conservation of important functional and structural residues points to the classical nature of the reported sea bass sequences. Nevertheless, insights on the polymorphism level and on the tissue distribution of individual genes, both hallmarks of classical MHC class I, are needed to confirm this. Finally all sea bass Dila-UA molecules were predicted to bind peptide ligands, according to the MHCLIG server (data not shown).

3.4. 3D homology modeling

The Dila-β2m and Dila-UA extracellular structures were predicted by comparative modeling using the grass carp β2m [PDB: 3gbl; (Chen et al., 2010b)] and chicken MHC [PDB: 3bev; (Koch et al., 2007)] crystal structures as templates, respectively (Fig. S4). The homology model of Dila-β2m is compatible with the seven-stranded β-sandwich fold typical of the IgSf constant (C) domain described for other β2m molecules. The seven β-strands are organized in two β-sheets with an ABDE-CFG arrangement stabilized by a disulfide between strands B and F (Fig. S4A). A conserved two-residue β-bulge ($hD^{53}L^{54}$/$sbD^{57}L^{58}$), which has been proposed to prevent protein aggregation (Richardson and Richardson, 2002) and to facilitate binding to the heavy chain (Trinh et al., 2002) is also present between the short D1 and D2 strands. Although the aspartate of this motif is widely conserved (glutamate in tongue sole), the leucine residue is replaced by methionine in mouse, chicken and tongue sole, and by proline in amphibians.

The homology model of Dila-UA is also structurally compatible with the PBD (α1 and α2) and the Ig fold (α3) described for other MHC class I HC molecules (Fig. S4B). As described for the crystal structures of human, mouse, chicken and bovine MHC class I, the modeled sea bass Dila-UA PBD is composed of two adjacent helical regions on top of an anti-parallel eight-stranded β-sheet (Figs. S4B and C), while the α3 domain consists of a typical Ig-fold (Fig. S4B). The eight anchor residues (red) important for peptide binding and

α2 domain

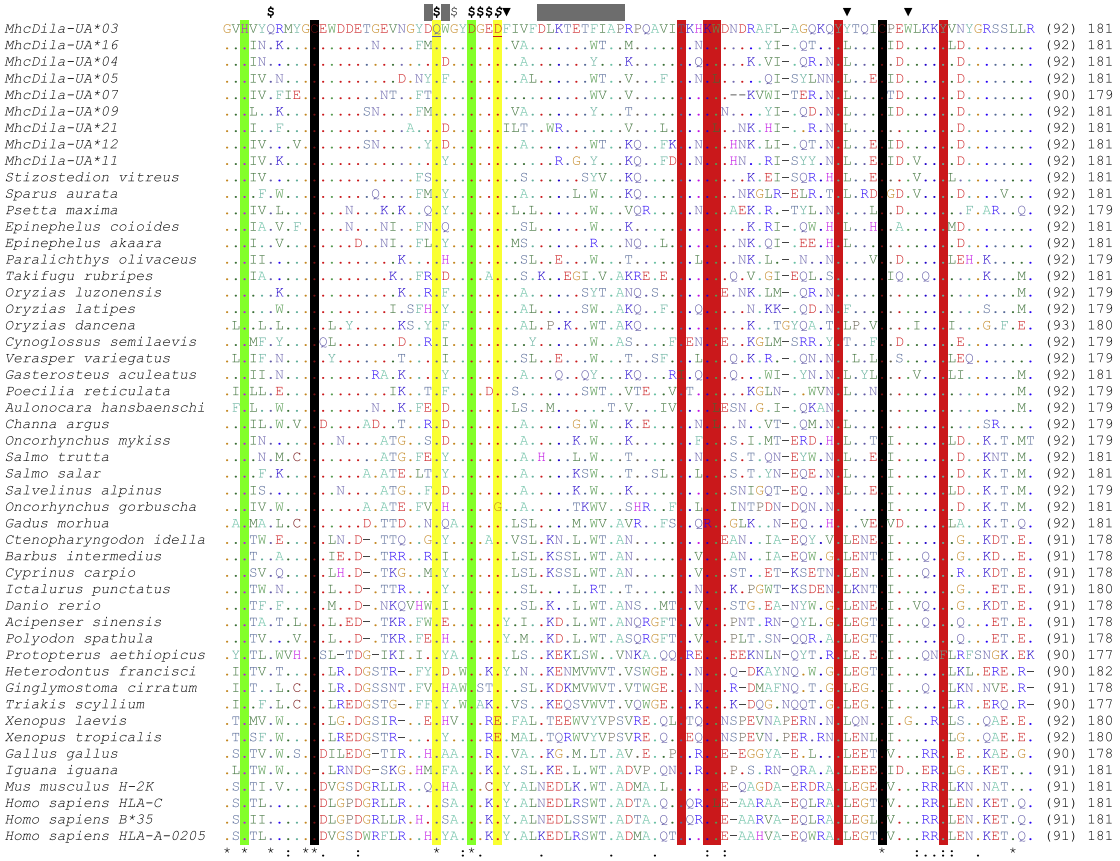


Fig. 5. (*continued*)

other highly conserved residues (pink) contacting the peptide occupy reasonable spatial positions (Fig. S4C). Three of the identified sea bass HC molecules (Dila-UA∗10-∗12) bear two additional cysteines within the α1 domain, which could form an extra disulfide bridge similar to that present in the α2 domain, linking the α1-helix to the floor of the groove (Fig. S4D). Despite one cannot discuss polymorphism given the undefined allele/locus nature of the transcripts, variable positions in the extracellular region were mapped to the homology model and as previously discussed highest diversity is located at the PBD (Fig. S4E). At last, the relative positions of sea bass β2m and HC in the context of the putative heterodimer, based on the chicken MHC class I structure are shown in Fig. S4I.

3.5. Phylogenetic analysis

In order to analyze the evolutionary relationships of the β2m and MHC class I HC proteins from different organisms and distinct lineages, a joint neighbor-joining tree was constructed (Fig. 6) with both groups of proteins clearly separated. As expected, Dila- β2m (mature protein) clusters with other bony fish sequences being closely related to molecules of species from the same order (e.g. Asian sea bass, yellow croaker and walleye). Sturgeon β2m clusters with the teleosts, although it represents a primitive form of Actinopterygii. Additionally, a series of pair-wise alignments were performed with MatGAT in order to calculate similarity and identity of β2m amino acid sequences between species. The results confirm those previously obtained with the neighbor-joining method (Table S3).

Regarding the MHC class I phylogeny (Fig. 6), all sea bass sequences (full-length) starkly diverge from MHC molecules of the L, S and Z teleost lineages (Dijkstra et al., 2007; Lukacs et al., 2010; Stet et al., 1998), clustering together and with other molecules of the U lineage (Stet et al., 1998), broadly represented among teleosts. Except for the Sarcopterygii molecule that falls within the Z lineage, all proteins are separated according to the class of the organisms: Mammalia, Reptilia, Aves, Amphibia, Chondrichthyes, and Actinopterygii. Within the Actinopterygii, it is obvious the separation between Teleostei and the ancient Chondrostei. As expected, the Dila-UA sequences are closely related to molecules of species from the Perciformes (e.g. *S. aurata*, *E. coioides*, *Epinephelus akaara* or *Stizostedion vitreus*) or Percomorpha (e.g. *G. aculeatus*). MatGAT analysis confirms these results (Table S4). These data suggest that all the sea bass clones belong to the teleost U lineage of MHC class I molecules.

α3 domain

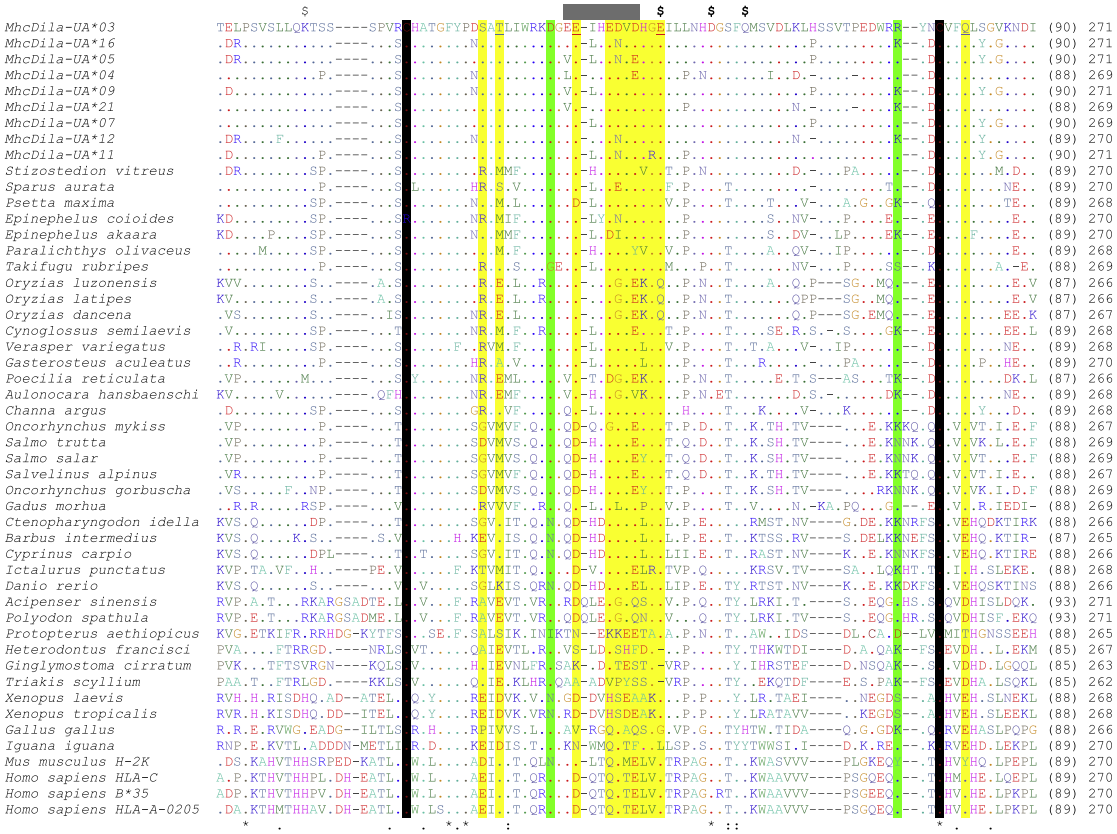


Fig. 5. (*continued*)

Finally, neighbor joining trees were constructed in order to evaluate the evolutionary relationships between sea bass class I molecules and those of other fish, separately analyzing the domains (Fig. 7). Here, the topologies of the α1, α2 and α3 domains differed, reflecting distinct evolutionary pathways of these functionally distinct regions (Fig. 7A–C). All teleost U lineage genes clustered together in all domains, except for α1, where some sequences group with Z lineage sequences. However, for this region the bootstrap values are not supportive of the observed topology (when pairwise deletion of gaps is used instead of all sites, Z lineage loci form a separate cluster with high bootstrapping also in α1 – not shown). Sea bass sequences of each extracellular region grouped altogether with high bootstrap values, but with low bootstrapping within the Dila-UA cluster (Fig. 7A–C), suggesting a recent expansion of sea bass class I molecules, similarly to the stickleback example (Sato et al., 1998). Here, repeated gene duplication and strong positive selection are hypothesized as the driving forces in the origin and maintenance of diversity in class Ia (Schaschl and Wegner, 2007). Recently, Nonaka et al. (2011) reported the existence of four across-species α1 lineages (L1–L4) established with high bootstrapping while studying medaka MHC class I. A previous study on rainbow trout UBA genes already attributed relevance to trans-species sharing of α1 lineages (Kiryu et al., 2005). The α1 domain tree generated with our data also displays these four lineages, suggesting all sea bass α1 domain sequences resemble those of the described L1 lineage (Fig. 7A). Similarly to the pufferfishes, they fit in the L1+ designation for having (i) two additional residues and (ii) in some sequences one or two cysteine residues (Table 1 and Fig. S3), features shared with L2 and L3 lineages (Nonaka et al., 2011). As a consequence of its long-term conservation, α1 yielded some insights on the orthologous relationships among teleost U lineage genes (Nonaka et al., 2011). Accordingly, L1 includes sequences from Acanthopterygii and salmonids (Protacanthopterygii), being considered the major α1 lineage of euteleostclass I HC genes (Nonaka et al., 2011). In the present data, no sea bass representatives are observed among the other domain lineages, in spite of L3 being also present in other Percomopha and Perciformes species and even in a sea bass partial sequence (Nonaka et al., 2011). Nevertheless, in accordance to the southern results, putatively undetected Dila-UA loci could belong to the mentioned L3 α1 lineage. The α2 tree (Fig. 7B) segregates the U lineage genes in two groups, in accordance to Nonaka et al. (2011) results, with all Dila-UA sequences included in the major one, although no further subdivisions could be inferred. Accumulation of point mutations and recombination events between sequences that increase their own diversity may be in the origin of such a topology (Nonaka et al., 2011). Regarding the α3 tree (Fig. 7C), species-specific (or related species) clusters were formed

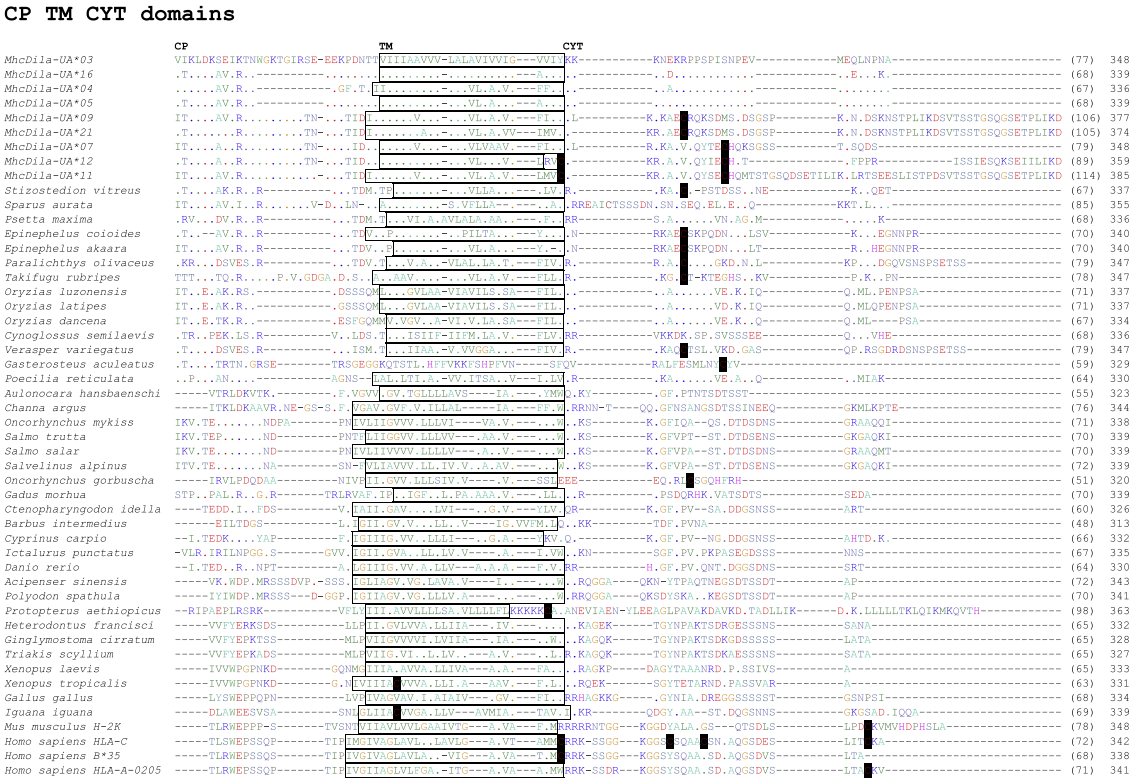


Fig. 5. (*continued*)

suggesting homogenization of α3 regions between loci after speciation, according to previous findings (Nonaka et al., 2011). Finally, two sub-lineages were recognized in sea bass for the CYT regions, designated Cyt-L1 and Cyt-L2 (Fig. 7D). Sub-lineage Cyt-L1 corresponds to the characteristically short tails, whilst the Cyt-L2 represents the longer ones. Similar results were obtained when the 3′ UTRs were used in the phylogenetic analysis (data not shown). Hence, based on the CYT and 3′ UTR regions, Dila-UA transcripts can be divided into two families, in contrast to medaka where inter-locus homogenization was observed for these domains (Nonaka and Nonaka, 2010). Locus specificities of TM-CYT and 3′ UTR regions were reported in human (Parham et al., 1989), mouse (Pullen et al., 1992) and Atlantic cod (Miller et al., 2002).

3.6. Basal expression analyses

The basal expression analysis of β2m in organs and tissues of non-stimulated sea bass is shown in Fig. 8A. Real-time PCR analysis detected the presence of β2m mRNA in all examined tissues demonstrating almost constitutive expression in juvenile sea bass, as observed in other teleosts (Choi et al., 2006; Shum et al., 1996; Yu et al., 2009). The highest level of β2m mRNA was detected in the thymus, the main site of T cell lymphopoiesis in vertebrate species (Hansen and Zapata, 1998) and the organ with the highest number of T-cells in sea bass (Abelli et al., 1996), followed by HK, PBL, spleen and gills with comparable levels; the lowest level of expression was detected in the muscle followed by brain. Comparing to other fish species, low β2m expression was found in the muscle and brain of walleye (*Stizostedium vtreum*) (Christie et al., 2007), large yellow croaker (*Pseudosciaena crocea*) (Yu et al., 2009) and rainbow trout (*Oncorhynchus mykiss*) (Shum et al., 1996), in agreement with the sea bass data, whereas the tissue with the highest expression is not conserved among the mentioned teleosts.

The MHC-I levels show a similar distribution when compared to those of β2m in the same tissues of sea bass, as shown in Fig. 8B, which is in agreement with the findings in rainbow trout (Shum et al., 1996). Hence, the highest expression was found in the thymus followed by gills, spleen, HK and gut, whilst the lowest expression was detected in the liver and brain (also in trout (Hansen et al., 1996)). These results of ubiquitous expression further support the classical nature of at least one or some of the Dila-UA sequences (Kaufman et al., 1994), since it was not determined which ones were amplified here.

3.7. β2m expression analyses after in vitro stimulation

To investigate whether β2m expression levels could be modulated with poly I:C, which simulates a viral infection, *in vitro* stimulation of HK leucocytes was studied both for short (6 h) and long (24 h) periods of time (Fig. 9). Head kidney was chosen, since in some teleosts it has been shown to have both the presence of diverse effector cell types in large numbers and of hematopoietic stem cells (HSCs) that in turn have the ability to reconstitute recipient lympho-hematopoietic organs, as in ginbuna carp (Kobayashi et al., 2006). A significant increase ($p < 0.05$) of β2m expression was detected after 24 h, whereas a small increase was evidenced after 6 h of stimulation. These results are in agreement with the up-regulation after poly I:C stimulation evidenced in large yellow croaker, *P. crocea* (Yu et al., 2009).

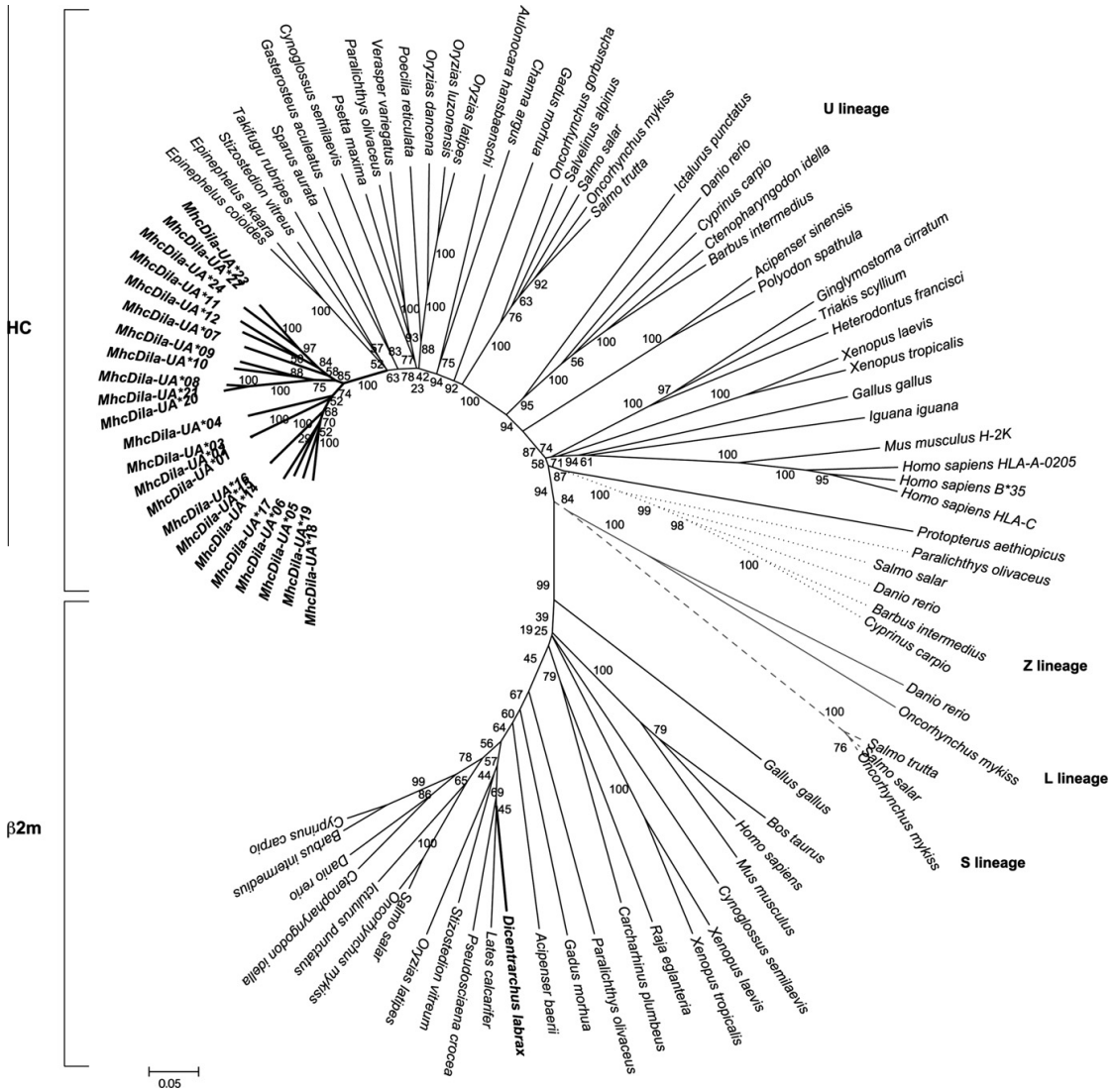


Fig. 6. Phylogenetic tree of β2m and MHC class I heavy chain amino acid sequences from different vertebrate species. neighbor-joining tree calculated with MEGA v5.05 (Tamura et al., 2011), using *p*-distance parameter and pairwise deletion of gaps. The branches were validated by bootstrap analysis from 1000 replications, represented by percentages in branch nodes. Mature and precursor forms were used for β2m and heavy chain sequences, respectively. These sequences are the same used in Figs. 4 and 5. Fish MHC class I molecules from distinct lineages (U, S, L and Z) were used and are labeled (accession numbers detailed in Table S2).

It has been demonstrated in bony fish that IFN- expression can be induced by poly I:C (Zou et al., 2005) and moreover, in mammals, INF- up-regulates the expression of MHC-class I heavy chain and β2m (Sugita et al., 1987), a pattern similar to the one detected in sea bass.

4. Conclusions

In the present study, the β2m cDNA and gene as well as several MHC class I HC transcripts and two partial genes from sea bass were cloned for the first time. All characteristic domains that are present in MHC class I complex light and heavy chains from other species are observed in both identified molecules. For β2m, characteristics such as exon/intron boundaries, gene number, tissue expression, and poly I:C inducibility (possible IFN-γ regulation) are also conserved, although the 3-exon structure is different from that of the human gene. For the HC, besides the PBD, striking variability was unexpectedly observed from the connective peptides through the 3′ UTR regions. Further genomic analysis of α1 through 3′ UTR regions in one individual fish revealed that two types of genes were being transcribed, containing distinct exon/intron structures and different exons and introns lengths. The observed specificities could potentially be responsible for distinct functions,

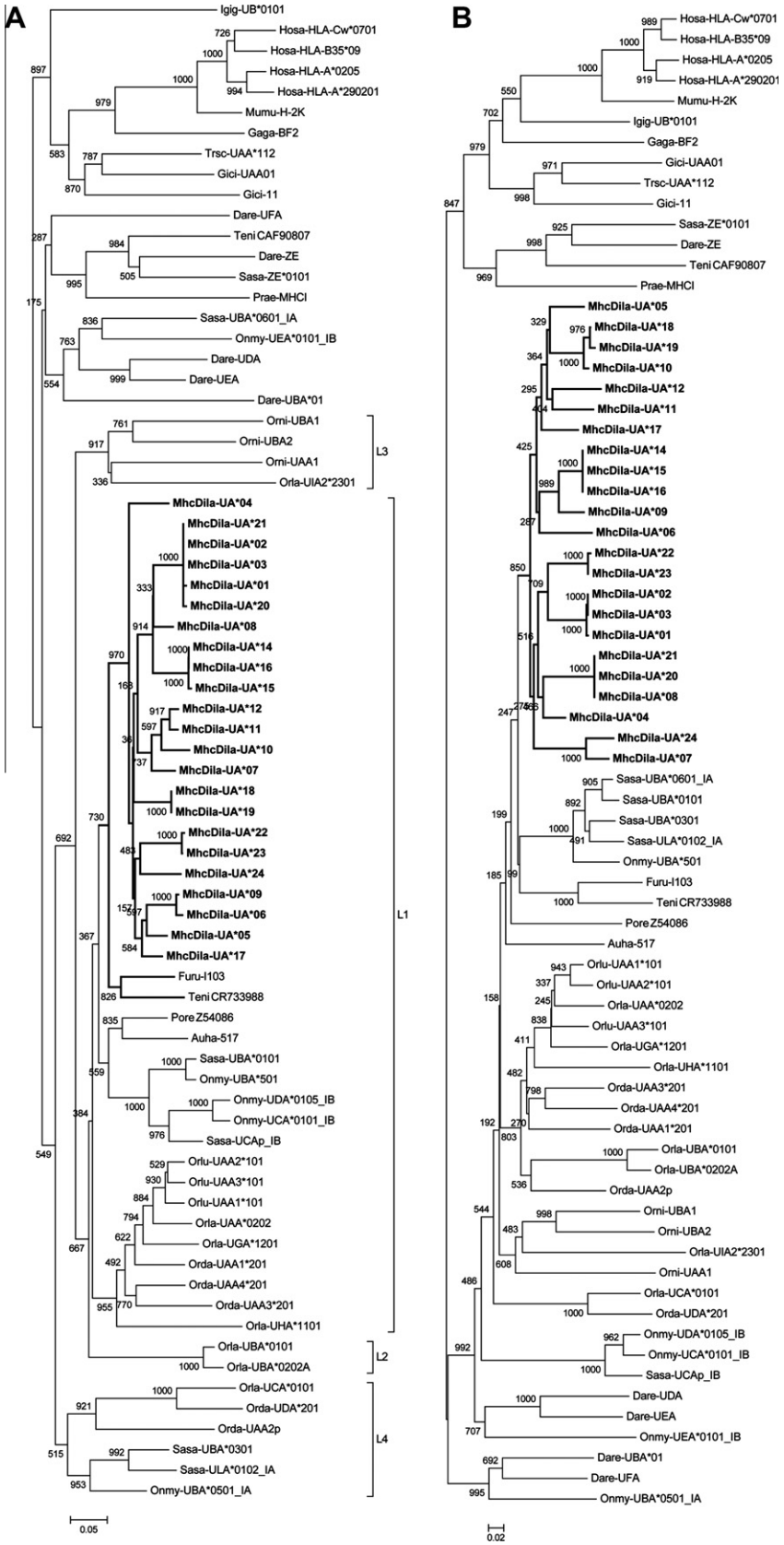


Fig. 7. Phylogenetic trees of MHC class I heavy chain separate domains from different vertebrate species. Nucleotide sequences were used to construct each neighbor-joining tree, calculated with ClustalX (Larkin et al., 2007) and edited with MEGA v5.05 (Tamura et al., 2011), using all sites. The branches were validated by bootstrap analysis from 1000 replications, represented by values in branch nodes. Sea bass sequences are in bold. Putative pseudogenes are indicated with suffix 'p'. Genes from the two MHC class I regions of the tetraploidsalmonid species are denoted with _IA or _IB. Accession numbers from sequences retrieved from NCBI GenBank are detailed on Table S2. (A) α1 domains. (B) α2 domains. (C) α3 domains. (D) Connective peptide, transmembrane and cytoplasmic regions. Medakaα1 domain L1–L4 lineages (Nonaka et al., 2011) as well as the two Dila-UA sub-lineages Cyt-L1 and Cyt-L2 are indicated. The three nile tilapia (Orni) sequences were not used in the cytoplasmic tails tree due to their extremely short length.

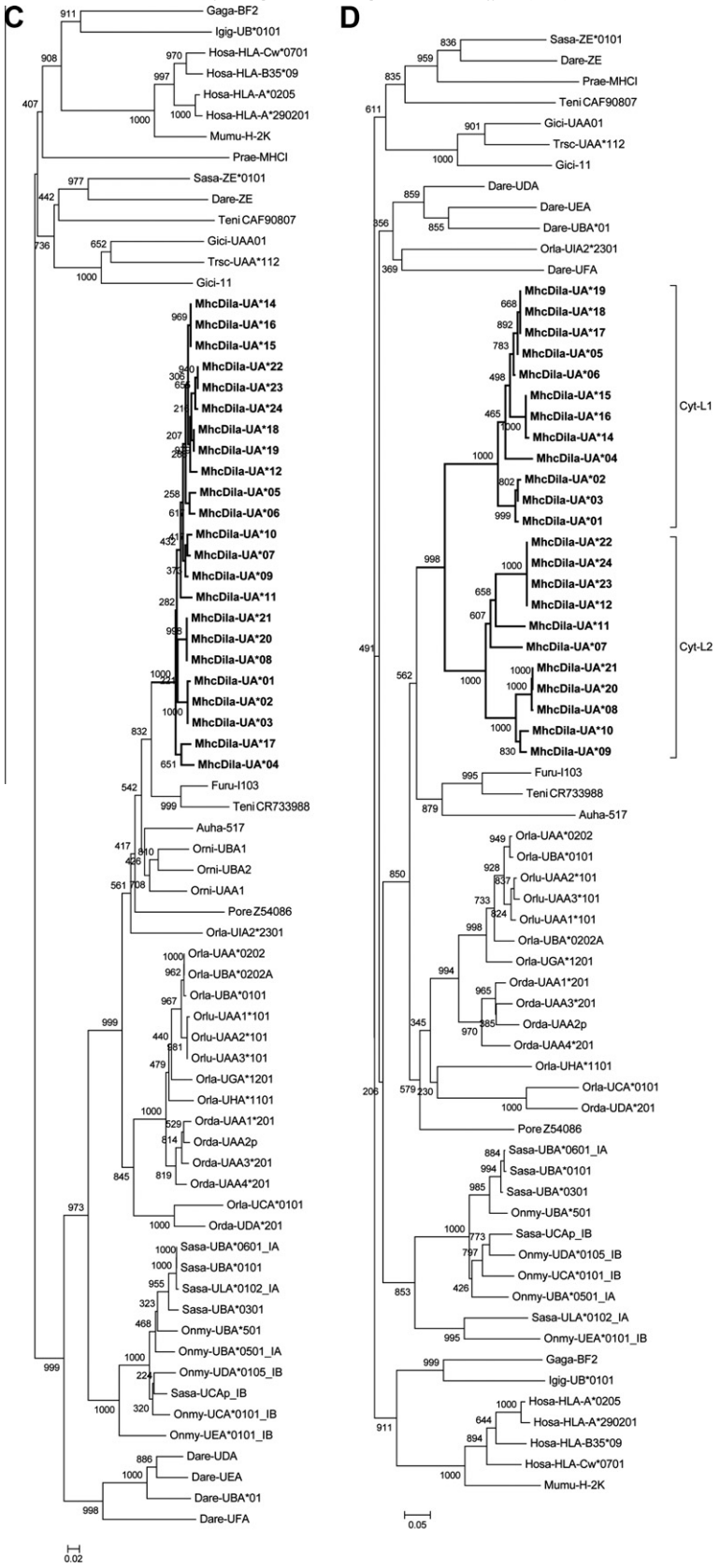


Fig. 7. (*continued*)

an issue that requires further elucidation. Nonetheless, sequence conservation of important residues, variability, polygeny and ubiquitous tissue-specific expression suggest that at least one or some of the sea bass clones here identified represent classical MHC class I genes. Phylogenetic analysis places all of them among the U lineage of teleost MHC class I α chains. Separate domain analysis

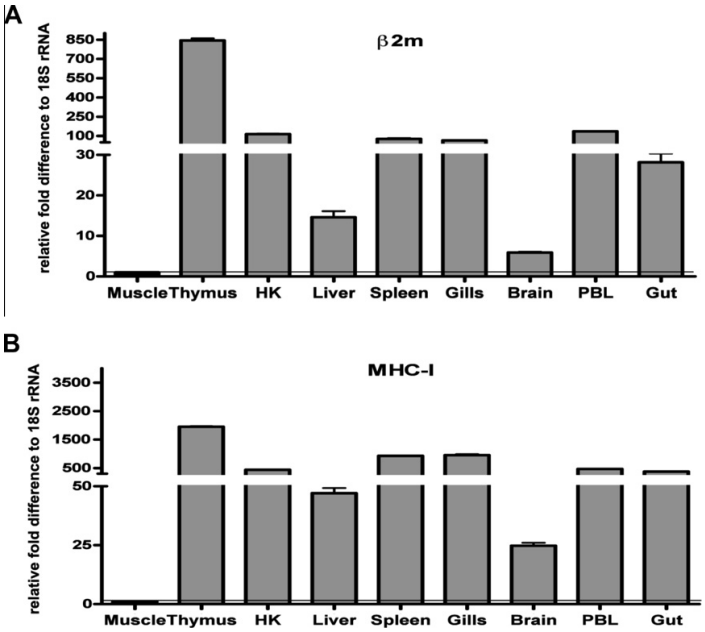


Fig. 8. Sea bass β2m and MHC class I heavy chain basal expression in different tissues. Both β2m (A)and MHC class I (B) mRNA levels were expressed as a ratio relative to rRNA 18S levels in the same samples after real-time PCR analysis using the tissue with the lowest expression (muscle) as calibrator.

classified all sea bass α1 regions as part of the trans-species L1 lineage, which includes Acanthopterygii and salmonid HC genes, being closely related to the pufferfish ones. Although no distinguishable Dila-specific sub-lineages among α1, α2 and α3 domains could be observed, two well-separated sub-lineages of CP/TM/CYT regions were identified (Cyt-L1 and Cyt-L2) possibly originated in different ancestral genes. Dila-β2m and Dila-UA 3D homology models have been compared to those of other vertebrate molecules. Altogether these results indicate that the reported Dila-β2m is orthologue of other β2m molecules while the Dila-UA transcripts are orthologues of other Acanthopterygii and salmonid U lineage MHC class I α genes, both providing important primary data for future studies on identification of virus-derived T-cell epitopes and other aspects of the antigen presentation pathway in this species.

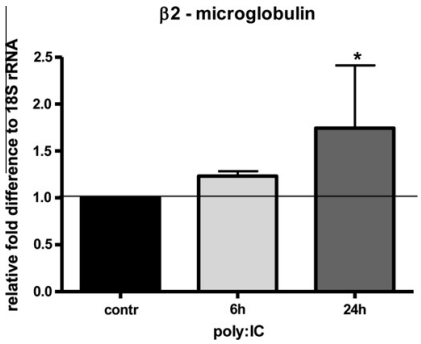


Fig. 9. Sea bass β2m expression after *in vitro* stimulation with poly I:C. β2m mRNA levels expressed as a ratio relative to rRNA 18S levels in the same samples after real-time PCR analysis of HK leucocytes stimulated with PBS (control) or with 50 g/ml of poly I:C for 6 and 24 h and normalized against the non-stimulated 0 h control. Controls for 4 and 24 h of incubation with PBS only are also shown in the graphs. Data were expressed as the mean ± SD and asterisks indicates when $p < 0.05$ with respect to the time 0 control.

Acknowledgments

We are grateful to the reviewers for helping to clarify and improving the manuscript. Funding from **FCT** – Fundação para a Ciência e a Tecnologia, Portugal (POCTI/CVT/44925/2006) supported this work. Rute D. Pinto acknowledges **FCT** PhD fellowship BD/42327/2007, funded by **POPH-QREN** and co-funded by **FSE** and **MCTES**. The funders had no role in study design, data collection and analysis, decision to publish, or publication of the manuscript. No additional external funding was received for this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dci.2012.10.002.

References

Abelli, L., Picchietti, S., Romano, N., Mastrolia, L., Scapigliati, G., 1996. Immunocytochemical detection of thymocyte antigenic determinants in developing organs of sea bass *Dicentrarchus labrax* (L.). Fish Shellfish. Immunol. 6, 493–505.

Achour, A., Persson, K., Harris, R.A., Sundback, J., Sentman, C.L., Lindqvist, Y., Schneider, G., Karre, K., 1998. The crystal structure of H-2Dd MHC class I complexed with the HIV-1-derived peptide P18–I10 at 2.4 A resolution: implications for T cell and NK cell recognition. Immunity 9, 199–208.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart,

P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D: Biol. Crystallogr. 66, 213–221.
Antoniou, A.N., Ford, S., Taurog, J.D., Butcher, G.W., Powis, S.J., 2004. Formation of HLA-B27 homodimers and their relationship to assembly kinetics. J. Biol. Chem. 279, 8895–8902.
Aoyagi, K., Dijkstra, J.M., Xia, C., Denda, I., Ototake, M., Hashimoto, K., Nakanishi, T., 2002. Classical MHC class I genes composed of highly divergent sequence lineages share a single locus in rainbow trout (*Oncorhynchus mykiss*). J. Immunol. 168, 260–273.
Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modeling. Bioinformatics 22, 195–201.
Barel, M.T., Pizzato, N., van Leeuwen, D., Bouteiller, P.L., Wiertz, E.J., Lenfant, F., 2003. Amino acid composition of alpha1/alpha2 domains and cytoplasmic tail of MHC class I molecules determine their susceptibility to human cytomegalovirus US11-mediated down-regulation. Eur. J. Immunol. 33, 1707–1716.
Batalia, M.A., Collins, E.J., 1997. Peptide binding by class I and class II MHC molecules. Biopolymers 43, 281–302.
Becker, J.W., Reeke Jr., G.N., 1985. Three-dimensional structure of beta 2-microglobulin. Proc. Natl. Acad. Sci. USA 82, 4225–4229.
Bendtsen, J.D., Nielsen, H., von Heijne, G., Brunak, S., 2004. Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340, 783–795.
Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343–350.
Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W.S., Johnson, S.L., Gates, M., Postlethwait, J.H., Klein, J., 1997. Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, *Danio rerio*. Immunogenetics 46, 129–134.
Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C., 1987a. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature 329, 512–518.
Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C., 1987b. Structure of the human class I histocompatibility antigen, HLA-A2. Nature 329, 506–512.
Buonocore, F.R.E., Bird, S., Secombes, C.J., Facchiano, A., Costantini, S., Scapigliati, G., 2007. Interleukin-10 expression by real-time PCR and homology modelling analysis in the European sea bass (*Dicentrarchus Labrax* L.). Aquaculture 270, 512–522.
Campanella, J.J., Bitincka, L., Smalley, J., 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 4, 29.
Chen, H., Kshirsagar, S., Jensen, I., Lau, K., Simonson, C., Schluter, S.F., 2010a. Characterization of arrangement and expression of the beta-2 microglobulin locus in the sandbar and nurse shark. Dev. Comp. Immunol. 34, 189–195.
Chen, M., Bouvier, M., 2007. Analysis of interactions in a tapasin/class I complex provides a mechanism for peptide selection. EMBO J. 26, 1681–1690.
Chen, W., Gao, F., Chu, F., Zhang, J., Gao, G.F., Xia, C., 2010b. Crystal structure of a bony fish beta2-microglobulin: insights into the evolutionary origin of immunoglobulin superfamily constant molecules. J. Biol. Chem. 285, 22505–22512.
Choi, W., Lee, E.Y., Choi, T.J., 2006. Cloning and sequence analysis of the beta2-microglobulin transcript from flounder, *Paralichthys olivaceous*. Mol. Immunol. 43, 1565–1572.
Christie, D., Wei, G., Fujiki, K., Dixon, B., 2007. Cloning and characterization of a cDNA encoding walleye (*Sander vitreum*) beta-2 microglobulin. Fish Shellfish. Immunol. 22, 727–733.
Criscitiello, M.F., Benedetto, R., Antao, A., Wilson, M.R., Chinchar, V.G., Miller, N.W., Clem, L.W., McConnell, T.J., 1998. Beta 2-microglobulin of ictalurid catfishes. Immunogenetics 48, 339–343.
Diedrich, G., Bangia, N., Pan, M., Cresswell, P., 2001. A role for calnexin in the assembly of the MHC class I loading complex in the endoplasmic reticulum. J. Immunol. 166, 1703–1709.
Dijkstra, J.M., Katagiri, T., Hosomichi, K., Yanagiya, K., Inoko, H., Ototake, M., Aoki, T., Hashimoto, K., Shiina, T., 2007. A third broad lineage of major histocompatibility complex (MHC) class I in teleost fish; MHC class II linkage and processed genes. Immunogenetics 59, 305–321.
Dixon, B., Stet, R.J., van Erp, S.H., Pohajdak, B., 1993. Characterization of beta 2-microglobulin transcripts from two teleost species. Immunogenetics 38, 27–34.
Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.D., Smith, J.A., Struhl, K., 1999. Current Protocols in Molecular Biology. John Wiley and Sons Inc., New York.
Figueroa, F., Mayer, W.E., Sato, A., Zaleska-Rutczynska, Z., Hess, B., Tichy, H., Klein, J., 2001. MHC class I genes of swordtail fishes, *Xiphophorus*: variation in the number of loci and existence of ancient gene families. Immunogenetics 53, 695–708.
Flajnik, M.F., Kasahara, M., 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. 11, 47–59.
Flajnik, M.F., Ohta, Y., Greenberg, A.S., Salter-Cid, L., Carrizosa, A., Du Pasquier, L., Kasahara, M., 1999. Two ancient allelic lineages at the single classical class I locus in the *Xenopus* MHC. J. Immunol. 163, 3826–3833.
Frickel, E.M., Riek, R., Jelesarov, I., Helenius, A., Wuthrich, K., Ellgaard, L., 2002. TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. Proc. Natl. Acad. Sci. USA 99, 1954–1959.
Gao, G.F., Tormo, J., Gerth, U.C., Wyer, J.R., McMichael, A.J., Stuart, D.I., Bell, J.I., Jones, E.Y., Jakobsen, B.K., 1997. Crystal structure of the complex between human CD8alpha(alpha) and HLA-A2. Nature 387, 630–634.
Goding, J.W., Walker, I.D., 1980. Allelic forms of beta 2-microglobulin in the mouse. Proc. Natl. Acad. Sci. USA 77, 7395–7399.
Grey, H.M., Kubo, R.T., Colon, S.M., Poulik, M.D., Cresswell, P., Springer, T., Turner, M., Strominger, J.L., 1973. The small subunit of HL-A antigens is beta 2-microglobulin. J. Exp. Med. 138, 1608–1612.
Grimholt, U., Drablos, F., Jorgensen, S.M., Hoyheim, B., Stet, R.J., 2002. The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics 54, 570–581.
Grimholt, U., Hordvik, I., Fosse, V.M., Olsaker, I., Endresen, C., Lie, O., 1993. Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). Immunogenetics 37, 469–473.
Gussow, D., Rein, R., Ginjaar, I., Hochstenbach, F., Seemann, G., Kottman, A., Ploegh, H.L., 1987. The human beta 2-microglobulin gene: primary structure and definition of the transcriptional unit. J. Immunol. 139, 3132–3138.
Halaby, D.M., Poupon, A., Mornon, J., 1999. The immunoglobulin fold family: sequence analysis and 3D structure comparisons. Protein Eng. 12, 563–571.
Hall, T., 1998. Bioedit. Biological Sequence Alignment Editor for Windows. North Carolina State University, NC, USA.
Hansen, J.D., Strassburger, P., Du Pasquier, L., 1996. Conservation of an alpha 2 domain within the teleostean world, MHC class I from the rainbow trout *Oncorhynchus mykiss*. Dev. Comp. Immunol. 20, 417–425.
Hansen, J.D., Strassburger, P., Thorgaard, G.H., Young, W.P., Du Pasquier, L., 1999. Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. J. Immunol. 163, 774–786.
Hansen, J.D., Zapata, A.G., 1998. Lymphocyte development in fish and amphibians. Immunol. Rev. 166, 199–220.
Hansen, T.H., Myers, N.B., Lee, D.R., 1988. Studies of two antigenic forms of Ld with disparate beta 2-microglobulin (beta 2m) associations suggest that beta 2m facilitate the folding of the alpha 1 and alpha 2 domains during de novo synthesis. J. Immunol. 140, 3522–3527.
Hao, H.F., Yang, T.Y., Yan, R.Q., Gao, F.S., Xia, C., 2006. CDNA cloning and genomic structure of grass carp (*Ctenophayngodon idellus*) beta2-microglobulin gene. Fish Shellfish. Immunol. 20, 118–123.
Higgins, D.G., 1994. CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol. Biol. 25, 307–318.
Hooft, R.W., Vriend, G., Sander, C., Abola, E.E., 1996. Errors in protein structures. Nature 381, 272.
Hutchinson, E.G., Thornton, J.M., 1996. PROMOTIF: a program to identify and analyze structural motifs in proteins. Protein Sci. 5, 212–220.
Kaufman, J., Milne, S., Gobel, T.W., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R., Beck, S., 1999. The chicken B locus is a minimal essential major histocompatibility complex. Nature 401, 923–925.
Kaufman, J., Salomonsen, J., Flajnik, M., 1994. Evolutionary conservation of MHC class I and class II molecules – different yet the same. Semin. Immunol. 6, 411–424.
Kelley, J., Walter, L., Trowsdale, J., 2005. Comparative genomics of major histocompatibility complexes. Immunogenetics 56, 683–695.
Kern, P.S., Teng, M.K., Smolyar, A., Liu, J.H., Liu, J., Hussey, R.E., Spoerl, R., Chang, H.C., Reinherz, E.L., Wang, J.H., 1998. Structural basis of CD8 coreceptor function revealed by crystallographic analysis of a murine CD8alphaalpha ectodomain fragment in complex with H-2Kb. Immunity 9, 519–530.
Kienast, A., Preuss, M., Winkler, M., Dick, T.P., 2007. Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin. Nat. Immunol. 8, 864–872.
Kimura, S., Tada, N., Yen, L.L., Hammerling, U., 1983. Exchange of cell-associated beta 2-microglobulin in mouse chimeras. Immunogenetics 18, 173–175.
Kiryu, I., Dijkstra, J.M., Sarder, R.I., Fujiwara, A., Yoshiura, Y., Ototake, M., 2005. New MHC class Ia domain lineages in rainbow trout (*Oncorhynchus mykiss*) which are shared with other fish species. Fish Shellfish. Immunol. 18, 243–254.
Klein, J., Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllensten, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K., Watkins, D.I., 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics 31, 217–219.
Kobayashi, I., Sekiya, M., Moritomo, T., Ototake, M., Nakanishi, T., 2006. Demonstration of hematopoietic stem cells in ginbuna carp (*Carassius auratus langsdorfii*) kidney. Dev. Comp. Immunol. 30, 1034–1046.
Koch, M., Camp, S., Collen, T., Avila, D., Salomonsen, J., Wallny, H.J., van Hateren, A., Hunt, L., Jacob, J.P., Johnston, F., Marston, D.A., Shaw, I., Dunbar, P.R., Cerundolo, V., Jones, E.Y., Kaufman, J., 2007. Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. Immunity 27, 885–899.
Kruiswijk, C.P., Hermsen, T., Fujiki, K., Dixon, B., Savelkoul, H.F., Stet, R.J., 2004. Analysis of genomic and expressed major histocompatibility class Ia and class II genes in a hexaploid Lake Tana African 'large' barb individual (*Barbus intermedius*). Immunogenetics 55, 770–781.
Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., Inoko, H., 2002. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol. Rev. 190, 95–122.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948.

Laskowski, R.A., MacArthur, M.W., Moss, D., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291.
Lizee, G., Basha, G., Tiong, J., Julien, J.P., Tian, M., Biron, K.E., Jefferies, W.A., 2003. Control of dendritic cell cross-presentation by the major histocompatibility complex class I cytoplasmic domain. Nat. Immunol. 4, 1065–1073.
Lukacs, M.F., Harstad, H., Bakke, H.G., Beetz-Sargent, M., McKinnel, L., Lubieniecki, K.P., Koop, B.F., Grimholt, U., 2010. Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. BMC Genomics 11, 154.
Lundqvist, M.L., Appelkvist, P., Hermsen, T., Pilstrom, L., Stet, R.J., 1999. Characterization of beta2-microglobulin in a primitive fish, the Siberian sturgeon (*Acipenser baeri*). Immunogenetics 50, 79–83.
Macdonald, I.K., Harkiolaki, M., Hunt, L., Connelley, T., Carroll, A.V., MacHugh, N.D., Graham, S.P., Jones, E.Y., Morrison, W.I., Flower, D.R., Ellis, S.A., 2010. MHC class I bound to an immunodominant *Theileria parva* epitope demonstrates unconventional presentation to T cell receptors. PLoS Pathog 6, e1001149.
Madden, D.R., 1995. The three-dimensional structure of peptide–MHC complexes. Annu. Rev. Immunol. 13, 587–622.
Magor, K.E., Shum, B.P., Parham, P., 2004. The beta 2-microglobulin locus of rainbow trout (*Oncorhynchus mykiss*) contains three polymorphic genes. J. Immunol. 172, 3635–3643.
Michalova, V., Murray, B.W., Sultmann, H., Klein, J., 2000. A contig map of the MHC class I genomic region in the zebrafish reveals ancient synteny. J. Immunol. 164, 5296–5305.
Miller, K.M., Kaukinen, K.H., Schulze, A.D., 2002. Expansion and contraction of major histocompatibility complex genes: a teleostean example. Immunogenetics 53, 941–963.
Miller, K.M., Li, S., Ming, T.J., Kaukinen, K.H., Schulze, A.D., 2006. The salmonid MHC class I: more ancient loci uncovered. Immunogenetics 58, 571–589.
Murray, B.W., Nilsson, P., Zaleska-Rutczynska, Z., Sultmann, H., Klein, J., 2000. Linkage relationships and haplotype variation of the major histocompatibility complex class I a genes in the cichlid fish *Oreochromis niloticus*. Mar. Biotechnol. (NY) 2, 437–448.
Nonaka, M.I., Aizawa, K., Mitani, H., Bannai, H.P., Nonaka, M., 2011. Retained orthologous relationships of the MHC class I genes during euteleost evolution. Mol. Biol. Evol. 28, 3099–3112.
Nonaka, M.I., Nonaka, M., 2010. Evolutionary analysis of two classical MHC class I loci of the medaka fish, *Oryzias latipes*: haplotype-specific genomic diversity, locus-specific polymorphisms, and interlocus homogenization. Immunogenetics 62, 319–332.
Ono, H., Figueroa, F., O'HUigin, C., Klein, J., 1993. Cloning of the beta 2-microglobulin gene in the zebrafish. Immunogenetics 38, 1–10.
Bjorkmann, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C., 1987. Structure of the human class I histocompatibility complex, HLA-A2. Nature 329, 506–512.
Parham, P., Lawlor, D.A., Lomen, C.E., Ennis, P.D., 1989. Diversity and diversification of HLA-A, B, C alleles. J. Immunol. 142, 3937–3950.
Parnes, J.R., Seidman, J.G., 1982. Structure of wild-type and mutant mouse beta 2-microglobulin genes. Cell 29, 661–669.
Peaper, D.R., Cresswell, P., 2008. Regulation of MHC class I assembly and peptide binding. Annu. Rev. Cell Dev. Biol. 24, 343–368.
Peaper, D.R., Wearsch, P.A., Cresswell, P., 2005. Tapasin and ERp57 form a stable disulfide-linked dimer within the MHC class I peptide-loading complex. EMBO J. 24, 3613–3623.
Persson, A.C., Stet, R.J., Pilstrom, L., 1999. Characterization of MHC class I and beta(2)-microglobulin sequences in Atlantic cod reveals an unusually high number of expressed class I genes. Immunogenetics 50, 49–59.
Pieters, J., Bakke, O., Dobberstein, B., 1993. The MHC class II-associated invariant chain contains two endosomal targeting signals within its cytoplasmic tail. J. Cell Sci. 106 (Pt 3), 831–846.
Praveen, P.V., Yaneva, R., Kalbacher, H., Springer, S., 2010. Tapasin edits peptides on MHC class I molecules by accelerating peptide exchange. Eur. J. Immunol. 40, 214–224.
Pullen, J.K., Horton, R.M., Cai, Z.L., Pease, L.R., 1992. Structural diversity of the classical H-2 genes: K, D, and L. J. Immunol. 148, 953–967.
Richardson, J.S., Richardson, D.C., 2002. Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. Proc. Natl. Acad. Sci. USA 99, 2754–2759.
Riegert, P., Andersen, R., Bumstead, N., Dohring, C., Dominguez-Steglich, M., Engberg, J., Salomonsen, J., Schmid, M., Schwager, J., Skjodt, K., Kaufman, J., 1996. The chicken beta 2-microglobulin gene is located on a non-major histocompatibility complex microchromosome: a small, G + C-rich gene with X and Y boxes in the promoter. Proc. Natl. Acad. Sci. USA 93, 1243–1248.
Rudolph, M.G., Stanfield, R.L., Wilson, I.A., 2006. How TCRs bind MHCs, peptides, and coreceptors. Annu. Rev. Immunol. 24, 419–466.
Sadasivan, B., Lehner, P.J., Ortmann, B., Spies, T., Cresswell, P., 1996. Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. Immunity 5, 103–114.
Saper, M.A., Bjorkman, P.J., Wiley, D.C., 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution. J. Mol. Biol. 219, 277–319.
Sarropoulou, E., Sepulcre, P., Poisa-Beiro, L., Mulero, V., Meseguer, J., Figueras, A., Novoa, B., Terzoglou, V., Reinhardt, R., Magoulas, A., Kotoulas, G., 2009. Profiling of infection specific mRNA transcripts of the European seabass *Dicentrarchus labrax*. BMC Genomics 10, 157.
Sato, A., Figueroa, F., Murray, B.W., Malaga-Trillo, E., Zaleska-Rutczynska, Z., Sultmann, H., Toyosawa, S., Wedekind, C., Steck, N., Klein, J., 2000. Nonlinkage of major histocompatibility complex class I and class II loci in bony fishes. Immunogenetics 51, 108–116.
Sato, A., Figueroa, F., O'HUigin, C., Steck, N., Klein, J., 1998. Cloning of major histocompatibility complex (Mhc) genes from threespine stickleback, *Gasterosteus aculeatus*. Mol. Mar. Biol. Biotechnol. 7, 221–231.
Scapigliati, G., Buonocore, F., Bird, S., Zou, J., Pelegrin, P., Falasca, C., Prugnoli, D., Secombes, C.J., 2001. Phylogeny of cytokines: molecular cloning and expression analysis of sea bass *Dicentrarchus labrax* interleukin-1beta. Fish Shellfish. Immunol. 11, 711–726.
Schaschl, H., Wegner, K.M., 2007. Contrasting mode of evolution between the MHC class I genomic region and class II region in the three-spined stickleback (*Gasterosteus aculeatus* L.; Gasterosteidae: Teleostei). Immunogenetics 59, 295–304.
Shaw, G., Kamen, R., 1986. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46, 659–667.
Shum, B.P., Azumi, K., Zhang, S., Kehrer, S.R., Raison, R.L., Detrich, H.W., Parham, P., 1996. Unexpected beta2-microglobulin sequence diversity in individual rainbow trout. Proc. Natl. Acad. Sci. USA 93, 2779–2784.
Shum, B.P., Guethlein, L., Flodin, L.R., Adkison, M.A., Hedrick, R.P., Nehring, R.B., Stet, R.J., Secombes, C., Parham, P., 2001. Modes of salmonid MHC class I and II evolution differ from the primate paradigm. J. Immunol. 166, 3297–3308.
Simister, N.E., Ahouse, J.C., 1996. The structure and evolution of FcRn. Res. Immunol. 147, 333–337, discussion 353.
Stet, R.J., Kruiswijk, C.P., Saeij, J.P., Wiegertjes, G.F., 1998. Major histocompatibility genes in cyprinid fishes: theory and practice. Immunol. Rev. 166, 301–316.
Stet, R.J., van Erp, S.H., Hermsen, T., Sultmann, H.A., Egberts, E., 1993. Polymorphism and estimation of the number of MhcCyca class I and class II genes in laboratory strains of the common carp (*Cyprinus carpio* L.). Dev. Comp. Immunol. 17, 141–156.
Stewart, R., Ohta, Y., Minter, R.R., Gibbons, T., Horton, T.L., Ritchie, P., Horton, J.D., Flajnik, M.F., Watson, M.D., 2005. Cloning and characterization of *Xenopus* beta2-microglobulin. Dev. Comp. Immunol. 29, 723–732.
Sugita, K., Miyazaki, J., Appella, E., Ozato, K., 1987. Interferons increase transcription of a major histocompatibility class I gene via a 5′ interferon consensus sequence. Mol. Cell Biol. 7, 2625–2630.
Sugita, M., Porcelli, S.A., Brenner, M.B., 1997. Assembly and retention of CD1b heavy chains in the endoplasmic reticulum. J. Immunol. 159, 2358–2365.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Trinh, C.H., Smith, D.P., Kalverda, A.P., Phillips, S.E., Radford, S.E., 2002. Crystal structure of monomeric human beta-2-microglobulin reveals clues to its amyloidogenic properties. Proc. Natl. Acad. Sci. USA 99, 9771–9776.
Tsukamoto, K., Hayashi, S., Matsuo, M.Y., Nonaka, M.I., Kondo, M., Shima, A., Asakawa, S., Shimizu, N., Nonaka, M., 2005. Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, *Oryzias latipes*. Immunogenetics 57, 420–431.
Turnquist, H.R., Vargas, S.E., Schenk, E.L., McIlhaney, M.M., Reber, A.J., Solheim, J.C., 2002. The interface between tapasin and MHC class I: identification of amino acid residues in both proteins that influence their interaction. Immunol. Res. 25, 261–269.
Tysoe-Calnon, V.A., Grundy, J.E., Perkins, S.J., 1991. Molecular comparisons of the beta 2-microglobulin-binding site in class I major-histocompatibility-complex alpha-chains and proteins of related sequences. Biochem. J. 277 (Pt 2), 359–369.
Van Hateren, A., James, E., Bailey, A., Phillips, A., Dalchau, N., Elliott, T., 2010. The cell biology of major histocompatibility complex class I assembly: towards a molecular understanding. Tissue Antigens 76, 259–275.
Vitiello, A., Potter, T.A., Sherman, L.A., 1990. The role of beta 2-microglobulin in peptide binding by class I molecules. Science 250, 1423–1426.
Wang, R., Natarajan, K., Margulies, D.H., 2009. Structural basis of the CD8 alpha beta/MHC class I interaction: focused recognition orients CD8 beta to a T cell proximal position. J. Immunol. 183, 2554–2564.
Wearsch, P.A., Cresswell, P., 2007. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. Nat. Immunol. 8, 873–881.
Wearsch, P.A., Cresswell, P., 2008. The quality control of MHC class I peptide loading. Curr. Opin. Cell. Biol. 20, 624–631.
Wearsch, P.A., Jakob, C.A., Vallin, A., Dwek, R.A., Rudd, P.M., Cresswell, P., 2004. Major histocompatibility complex class I molecules expressed with monoglucosylated N-linked glycans bind calreticulin independently of their assembly status. J. Biol. Chem. 279, 25112–25121.
Williams, A.F., Barclay, A.N., 1988. The immunoglobulin superfamily – domains for cell surface recognition. Annu. Rev. Immunol. 6, 381–405.
Xu, T.J., Sha, Z.X., Chen, S.L., 2010. Unexpected variations of beta2-microglobulin gene in the half-smooth tongue sole. Fish Shellfish. Immunol. 28, 212–215.
Yu, S., Chen, X., Ao, J., 2009. Molecular characterization and expression analysis of beta2-microglobulin in large yellow croaker *Pseudosciaena crocea*. Mol. Biol. Rep. 36, 1715–1723.
Zdobnov, E.M., Apweiler, R., 2001. InterProScan: an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.
Zou, J., Carrington, A., Collet, B., Dijkstra, J.M., Yoshiura, Y., Bols, N., Secombes, C., 2005. Identification and bioactivities of IFN-gamma in rainbow trout *Oncorhynchus mykiss*: the first Th1-type cytokine characterized functionally in fish. J. Immunol. 175, 2484–2494.

Supplementary data

Table S1

Primers used in this study.

Designation	Nucleotide sequence 5'-3'	Usage
AAP	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	5'RACE
APv	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTV	cDNA synthesis; PCR (full cDNA)
AUAP	GGCCACGCGTCGACTAGTAC	5'RACE
APv2	GACTCAGGACTTCAGGACTTAGTTTTTTTTTTTTTTTTV	cDNA synthesis; PCR (full cDNA)
AUAP2	GACTCAGGACTTCAGGACTTAG	PCR (full cDNA)
T7	TAATACGACTCACTATAGGGCGA	sequencing
SP6	CTATTTAGGTGACACTATAGAATAC	sequencing
b2mFW1	CCHCCHAARRTSSAIGTITA	PCR (partial cDNA)
b2mRV3	TCRAARGCSARRTCIGTYTG	PCR (partial cDNA)
DLb2mFW1	TAGTCGCTCCTGTTTTTCAT	PCR (full cDNA and gene)
DLb2mFW3	GCCGGAGGGATGAACATTAT	sequencing (full gene)
DLb2mFW4	CTAAACCCAAGGTTCAGGTGTA	southern blot
B2M FW	GGTTCAGGTGTACAGCC	qPCR
DLb2mRV1	AAGGCGAGGTCGGTCTGCTT	5'RACE; cDNA synthesis
DLb2mRV2	ATCCTTCAGCAACTGGATGT	5'RACE
DLb2mRV4	AAATTTAAAGAGGCATTTAT	PCR (full gene)
DLb2mRV6	GCACCCCCACATTCTCTTAC	sequencing (full gene)
DLb2mRV7	CCCAGGCATAGGTTTTCGTTTC	southern blot
B2M RV	CCTTGTGGGTGTGAAGG	qPCR
MHCIFW1	GWSMYCRAYTTYCCIGARTT	PCR (partial cDNA)
MHCIRV1	TCRAAICCIGTIGCRTGRCA	PCR (partial cDNA)
DLMHCIFW2	GGAGCTCTCTGCTGAGAACA	sequencing (full cDNA)
DLMHCIFW3	CAACTGGAAACAGTGTGTGTGAG	PCR (full cDNA)
DLMHCIFW4	GTGTGTGAGTTTCTCTGGATAAAG	PCR (full cDNA)
DLMHCIFW5	AGGCATCACAGCACTTTCTA	sequencing (full cDNA)
DLMHCIFW8	GCTGCCACGCTACAGGTTTC	PCR (partial gene)
DLMHCIFW18	CGATGGATCCTTCCAGATGAG	PCR (partial gene)
MHC-I FW	AARMAMTACCTCACCCAGR	qPCR
DLMHCIRV1	GCCGTACATGTTCTGGACAA	5'RACE; cDNA synthesis
DLMHCIRV2	CGGCTGTGACTCTGCTCATC	5'RACE
DLMHCIRV4a	GTTGAACAGTTTCTGTAAAT	sequencing (full cDNA)
DLMHCIRV4b	CTGCATGTAAATCACATAAC	sequencing (full cDNA)
DLMHCIRV4c	TCAACATTCTCGTCACTCTT	sequencing (full cDNA)
DLMHCIRV5	ACCTCTTCAATCTGTCACTC	sequencing (full cDNA)
DLMHCIRV8	CACCATAGAGCTGAAACACACAGTC	southern blot
DLMCHIRV14	CTCAGAAGTTCAATGTTGCAGATTT	PCR (partial gene, Cyt-L2)
DLMHCIRV15	AGTCCAAGTGAACTACACATCTCATT	PCR (full cDNA, Cyt-L2)
DLMHCIRV17	ACAGGACTTTATTGAAATCAAAAGTACA	PCR (full cDNA, Cyt-L1)
DLMHCIRV22	CGCTCCTCGTTCGAGTGCAAAA	PCR (partial gene, Cyt-L1)
MHC-I RV	CTCCATCTTTCCTCCAGAT	qPCR
RIB FW	CCAACGAGCTGCTGACC	qPCR
RIB RV	CCGTTACCCGTGGTCC	qPCR

Letter nucleotide code: H= A, C, T; R= A, G; Y= C, T; S= G, C; W= A, T; V= A, G, C; N= A, T, C, G; I= deoxyinosine.

Table S2

Accession numbers of sequences used in this study.

Species	β2m	MHC-I HC U lineage	Z lineage	L lineage	S lineage	Separate domains
yellow croaker (*P. crocea*)	ABB60035	-	-	-	-	-
walleye (*S. vitreum*)	AAW65850	AAL11412	-	-	-	-
Asian sea bass (*L. calcarifer*)	ACN32201	-	-	-	-	-
Japanese medaka (*O. latipes*)	NP_001098130	NP_001098573	-	-	-	Orla-UAA (*0202, AB450991), Orla-UBA (*0101, AB183488), Orla-UCA(*0101, AB033381), Orla-UDA (*0202, AB450998), Orla-UGA (*1201, AB604102), Orla-UHA (*1101, AB604112), Orla-UIA (2*2301. AB604119)
Atlantic salmon (*S. salar*)	AAG17535	AAN75110	AAZ76730	-	ACY30362	Sasa-UBA*0601_IA (AF504013), Sasa-UBA*0101 (AF504019), Sasa-UBA*0301 (AF504022), Sasa-ULA_IA (EF441211), Sasa-UCAp_IB (EF427379), Sasa-ZE (DQ099914)
rainbow trout (*O. mykiss*)	AAB04657	AAG25203	-	ABI21844	AAG27480	Onmy-UBA*0501_IA (AB162342), Onmy-UBA*501 (AF287488), Onmy-UCA_IB, Onmy-UDA_IB, and Onmy-UEA_IB (AB162343)
channel catfish (*I. punctatus*)	O42197	AAD08649	-	-	-	-
grass carp (*C. idella*)	BAD51478	ABU86895	-	-	-	-
zebrafish (*D. rerio*)	AAA50022	AAH74095	CAD12790	CAD56801	-	Dare-UBA01 (Z46777), Dare-UDA and Dare-UFA(AL672151), Dare-UEA (BC097061), Dare-ZE (NM_194425)
African barb (*B. intermedius*)	CAD44965	CAD44956	CAD12383	-	-	-
common carp (*C. carpio*)	Q03422	CAA62497	CAD12793	-	-	-
sturgeon (*A. baerii*)	CAB61324	-	-	-	-	-
Atlantic cod (*G. morhua*)	CAA10761	AAL14532	-	-	-	-
Japanese flounder (*P. olivaceus*)	AAN40738	BAD13367	BAD13366	-	-	-
tongue sole (*C. semilaevis*)	ACR38924	ACI88824	-	-	-	-
gilthead seabream (*S. aurata*)	-	ABB04088	-	-	-	-
orange-spotted grouper (*E. coioides*)	-	ACZ97571	-	-	-	-
Hong-Kong grouper (*E. akaara*)	-	ABX80523	-	-	-	-
fort maguire (*A. hansbaenschi*)	-	AAD37815	-	-	-	Auha-517 (AF038550)
snakehead (*C. argus*)	-	ACF74832	-	-	-	-
medaka (*O. luzunensis*)	-	ACN49175	-	-	-	Orlu-UAA1, Orlu-UAA2, and Orlu-UAA (FJ481085)
medaka (*O. dancena*)	-	ACN53465	-	-	-	Orda-UAA1, Orda-UAA2, Orda-UAA3, Orda-UAA4,and Orda-UDA (FJ481082)
turbot (*P. maxima*)	-	ABM92962	-	-	-	-
three-spinned stickleback (*G. aculeatus*)	-	ABN14356	-	-	-	-
spotted halibut (*V. variegatus*)	-	ADK26454	-	-	-	-
Japanese pufferfish (*T. rubripes*)	-	AAC41238	-	-	-	Furu-I103 (AJ271723)
pufferfish (*T. nigroviridis*)	-	-	-	-	-	Teni-CR733988 (CR733988), Teni-CAF90807(ZE) (CAAE01007789)
guppy (*P. reticulata*)	-	CAA90791	-	-	-	Pore-CAA90792 (Z54086)
Artic char (*S. alpinus*)	-	ACI05077	-	-	-	-
brown trout (*S. trutta*)	-	AAG02523	-	-	AAG09925	-
pink salmon (*O. gorbuscha*)	-	BAA09553	-	-	-	-
Chinese sturgeon (*A. sinensis*)	-	ACV87437	-	-	-	-
paddlefish (*P. spathula*)	-	ACV87421	-	-	-	-
marbled lungfish (*P. aethiopicus*)	-	AAF15304	-	-	-	Prae-AAF15304 (AF206309)
Nile tilapia (*O. niloticus*)	-	-	-	-	-	Orni-UAA1 , Omi-UBA1, and Orni-UBA2 (AB270897)
skate (*R. eglanteria*)	AAN62852	-	-	-	-	-
sandbar shark (*C. plumbeus*)	ACX46636	-	-	-	-	-
horn shark (*H. francisci*)	-	AAC60349	-	-	-	-
banded houndshark (*T. scyllium*)	-	AAB97335	-	-	-	Trsc-UAA (AF034327)
nurse shark (*G. cirratum*)	-	AAF66110	-	-	-	Gici-UAA01 (AF220063), Gici-11_UCA (AF028557)
Atlantic clawed frog (*X. laevis*)	AAF37230	AAA16064	-	-	-	-
western clawed frog (*X. tropicalis*)	BQ389924	NP_001136291	-	-	-	-
chicken (*G. gallus*)	P21611	NP_001026509	-	-	-	Gaga-BF2 (B21, AB426152)
iguana (*I. iguana*)	-	ACD54837	-	-	-	ACD54837
mouse, (*M. musculus*)	P01887	AAA39553	-	-	-	AAA39553
cow (*B. taurus*)	NP_776318	-	-	-	-	-
human (*H. sapiens*)	P61769	BAA22206; P30685; AAA03603	-	-	-	Hosa-A*290901 (AM944568), Hosa-Cw*0701 (BAA22206), Hosa-B*35 (P30685), Hosa-A*0205 (AAA03603)

Table S3

Percentages of amino acid similarity and identity among mature β2m molecules from different species.

	Similarity (%)	Identity (%)
1. Dicentrarchus labrax	-	-
2. Pseudosciaena crocea	82.2	73.3
3. Stizostedion vitreum	83.2	71.3
4. Lates calcarifer	80.2	74.3
5. Oryzias latipes	79.2	63.4
6. Salmo salar	81.2	67.3
7. Oncorhynchus mykiss	80.2	65.3
8. Ictulurus punctatus	81.2	64.4
9. Ctenopharyngodon idella	79.2	59.4
10. Danio rerio	81.2	59.4
11. Barbus intermedius	79.2	63.4
12. Cyprinus carpio	75.2	62.4
13. Acipenser baerii	75.2	61.2
14. Gadus morhua	68.3	53.5
15. Paralichthys olivaceus	75.2	52.5
16. Cynoglossus semilaevis	66.3	45.5
17. Raja eglanteria	58.4	39.6
18. Carcharhinus plumbeus	59.4	44.6
19. Xenopus laevis	62.4	37.3
20. Xenopus tropicalis	54.5	35.3
21. Gallus gallus	58.4	41.7
22. Mus musculus	66.3	43.3
23. Bos taurus	65.3	44.7
24. Homo sapiens	63.4	43.7

Table S4

Percentages of similarity and identity among MHC class heavy chain molecules from different species.

	Dila-UA Cyt-L1		*Dlia-UA Cyt-L2*	
	Similarity (%)	**Identity (%)**	**Similarity (%)**	**Identity (%)**
Dila-UA Cyt-L1	84.5-100	78.7-99.7	74.3-84.2	66.6-77.5
Dlia-UA Cyt-L2	74.3-84.2	66.6-77.5	79.4-99.5	71.0-99.7
Stizostedion vitreus	81.5-86.3	69.6-**73.5**	72.7-79.3	62.3-67.0
Sparus aurata	76.2-79.9	66.3-69.0	75.1-78.3	64.2-66.7
Psetta maxima	77.9-80.8	66.0-68.5	68.2-74.1	57.5-62.7
Epinephelus coioides	77.7-81.3	65.5-67.8	71.7-74.7	59.4-63.3
Epinephelus akaara	78.5-81.9	64.9-67.5	73.3-77.1	59.7-64.3
Paralichthys olivaceus	77.3-80.3	63.0-66.7	71.5-74.8	58.3-62.9
Takifugu rubripes	71.9-76.0	60.3-64.1	68.7-74.7	56.9-62.5
Oryzias luzonensis	76.5-80.4	61.0-65.2	68.1-73.0	53.2-58.5
Oryzias latipes	76.6-80.3	60.3-66.3	68.1-73.5	54.0-58.1
Oryzias dancena	72.8-77.0	59.0-62.9	64.6-71.4	50.6-55.7
Gasterosteus aculeatus	70.1-75.2	58.3-63.3	65.9-71.4	54.1-58.3
Poecilia reticulata	70.8-74.1	58.4-62.5	62.6-68.4	48.5-55.2
Cynoglossus semilaevis	74.9-79.2	58.2-61.9	67.6-73.0	52.7-56.8
Verasper variegatus	73.4-76.4	57.3-61.9	67.9-72.2	53.2-57.6
Aulonocara hansbaenschi	72.1-76.3	57.1-59.0	64.1-72.5	50.0-56.2
Channa argus	68.9-73.6	54.5-59.5	65.1-69.3	51.3-55.1
Oncorhynchus mykiss	72.2-76.5	55.8-59.7	64.7-72.5	50.5-56.6
Salmo trutta	68.9-72.5	54.3-56.7	63.1-68.1	49.0-54.8
Salmo salar	70.9-73.2	53.1-55.3	63.4-70.6	47.7-53.3
Salvelinus alpinus	72.5-75.1	53.0-55.2	66.3-73.0	49.3-54.3
Oncorhynchus gorbuscha	65.4-68.7	50.3-53.9	59.5-65.9	45.9-51.6
Gadus morhua	68.7-70.9	51.7-55.4	62.9-68.7	49.4-55.3
Ctenopharyngodon idella	62.1-65.4	44.9-47.1	56.6-62.1	41.9-44.3
Barbus intermedius	60.1-62.8	42.2-45.9	54.2-59.9	38.2-42.5
Cyprinus carpio	59.7-64.0	43.8-46.8	54.0-61.9	40.1-44.8
Ictalurus punctatus	59.8-61.7	42.9-45.9	54.7-61.9	39.2-46.0
Danio rerio	59.4-61.7	39.1-42.1	54.5-59.9	36.9-41.4
Acipenser sinensis	58.3-62.4	39.9-42.5	53.4-61.3	36.4-41.7
Polyodon spathula	57.4-60.4	39.5-41.9	53.2-61.6	36.5-42.7
Protopterus aethiopicus	46.4-48.7	**27.0**-30.2	48.2-52.3	27.4-29.2
Heterodontus francisci	54.8-56.7	36.7-39.5	50.6-55.6	33.2-36.4
Triakis scyllium	53.9-57.0	36.2-39.4	49.8-54.2	33.3-36.7
Ginglymostoma cirratum	51.2-55.2	35.5-38.6	48.4-52.3	32.8-36.0
Xenopus laevis	52.2-56.3	33.3-36.4	49.6-54.6	31.8-35.8
Xenopus tropicalis	53.1-56.7	34.6-36.7	50.7-56.4	31.5-35.2
Gallus gallus	52.9-55.0	31.9-33.9	48.0-54.0	29.5-35.0
Iguana iguana	51.0-55.1	32.3-35.4	47.8-51.7	29.6-33.2
Mus musculus H-2K	52.6-54.7	31.9-33.7	48.1-52.3	29.6-32.8
Homo sapiens HLA-C	49.7-51.6	28.6-30.6	47.6-50.7	26.7-30.6
Homo sapiens B*35	51.8-54.4	28.6-30.2	49.1-52.4	28.3-30.3
Homo sapiens HLA-A-0205	50.7-53.2	28.8-29.8	47.9-51.3	27.5-29.8
Barbus intermedius ZE	43.9-48.2	27.0-29.9	47.0-50.4	27.8-30.6
Cyprinus carpio ZE	45.9-48.0	28.3-31.7	47.5-50.3	28.4-30.3
Danio rerio ZE	46.7-49.7	28.1-32.2	48.2-50.7	28.2-32.1
Salmo salar ZE	47.7-50.3	28.4-31.0	43.1-49.3	25.9-30.8
Paralichthys olivaceus ZE	43.3-46.3	26.2-27.3	39.8-44.4	23.3-27.1
Danio rerio L	43.0-45.9	24.7-26.1	40.1-44.4	23.2-26.7
Oncorhynchus mykiss LBA	46.2-49.6	24.9-26.8	44.1-47.5	23.4-26.4
Salmo salar SAA	45.2-48.6	23.1-25.2	40.3-45.2	19.8-23.0
Oncorhynchus mykiss SAA	45.8-49.4	23.3-25.2	41.8-45.5	20.0-22.9
Salmo trutta SAA	43.3-46.4	22.3-24.7	39.6-43.1	19.8-21.7

Figure S1. Alignment of MHC class I heavy chain nucleotide sequences with short cytoplasmic tails. Nucleotide sequences were aligned with CLUSTALW. Dashes indicate gaps that maximize the alignment, and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with, (*), (:) and (.), respectively. Numbers right to the sequences denote their lengths. 5'UTRs, distinct domains and 3'UTRs are indicated above the alignment. Start and stop codons are bold and underlined. Grey shading highlights a different start of domain. Cyan shading denotes location of intron according to MhcDila-UA*25.

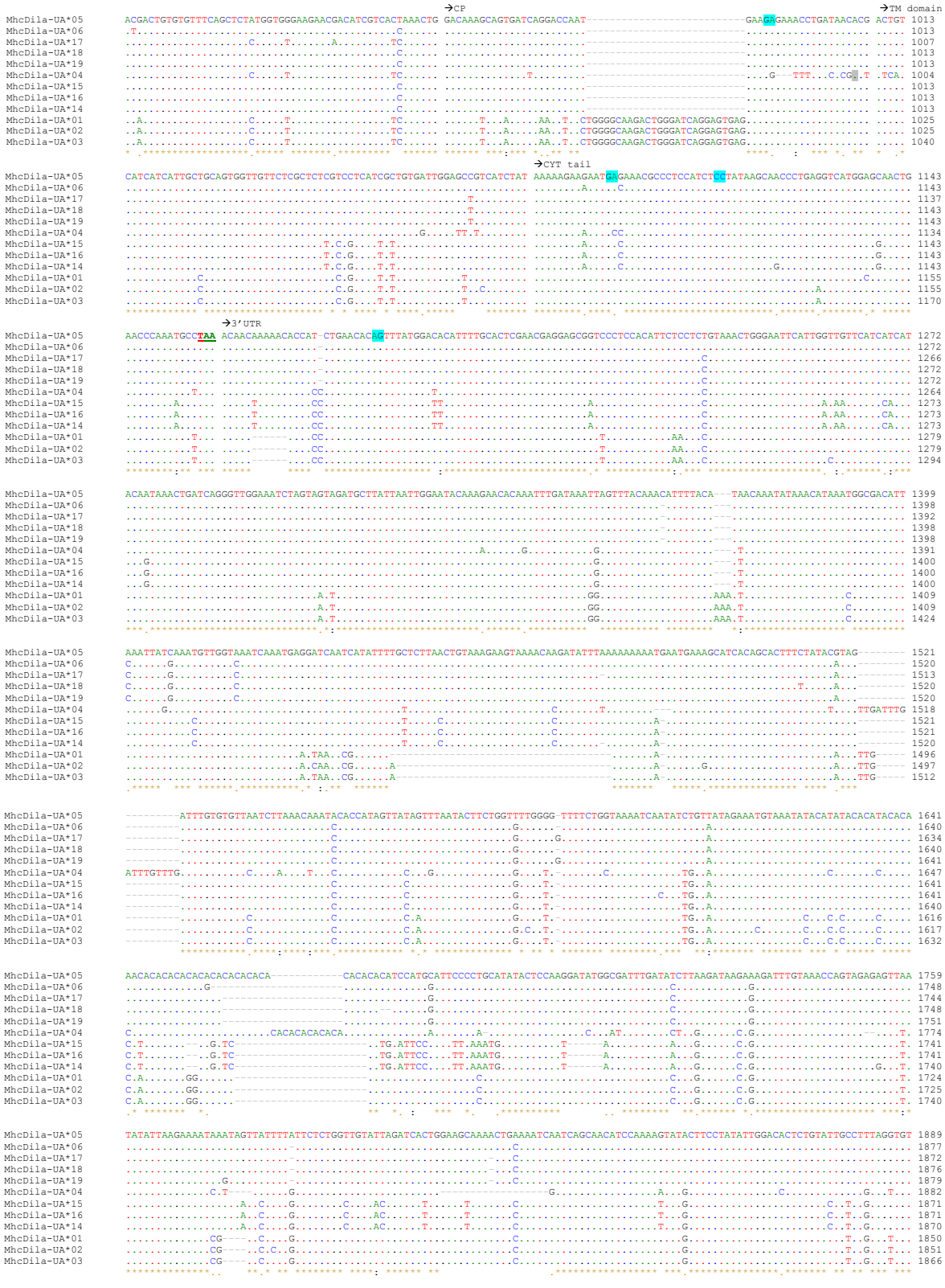


Figure S1 *(continued)*

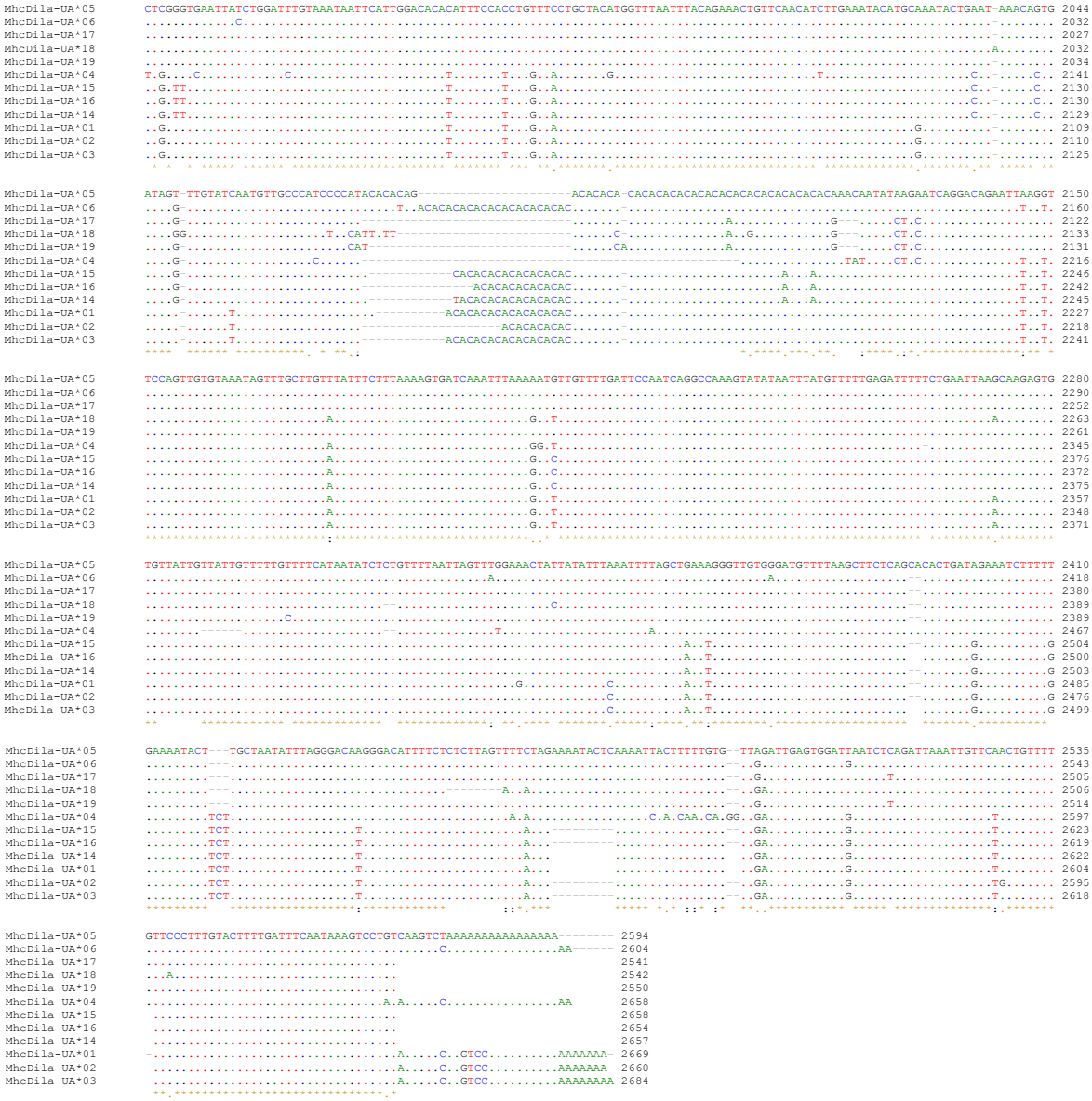


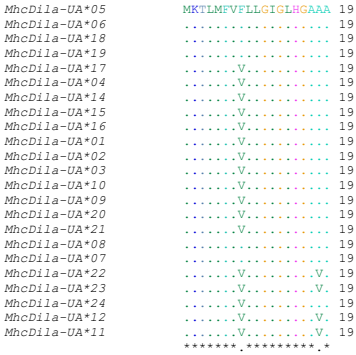
Figure S1 *(continued)*

Figure S2. Alignment of MHC class I heavy chain nucleotide sequences with long cytoplasmic tails. Nucleotide sequences were aligned with CLUSTALW. Dashes indicate gaps that maximize the alignment, and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with, (*), (:) and (.), respectively. Numbers right to the sequences denote their lengths. 5' UTRs, distinct domains and 3' UTRs are indicated above the alignment. Start and stop codons are bold and underlined. Grey shading highlights a different start of domain. Cyan shading denotes location of intron according to MhcDila-UA*26.

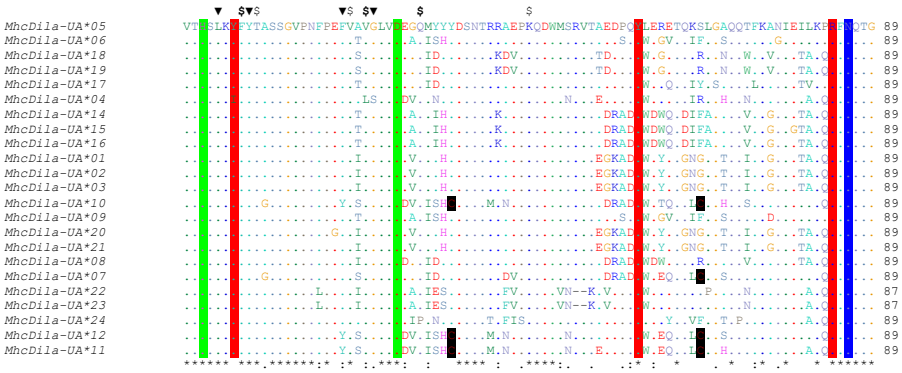
Figure S2 *(continued)*

Figure S2 *(continued)*

Signal Peptide



α1 domain



α2 domain

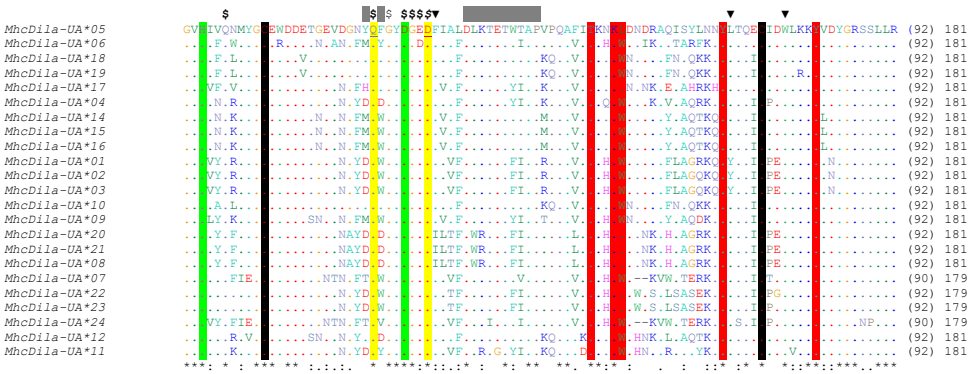
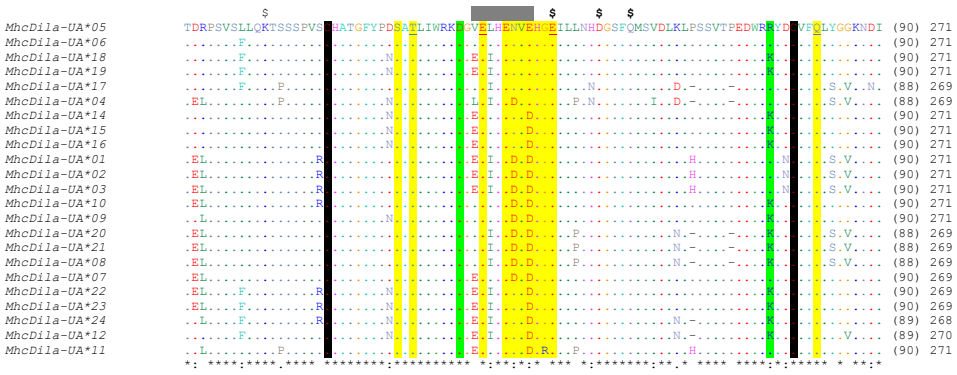


Figure S3. Alignment of sea bass MHC class I heavy chain amino acid sequences. Amino acid sequences were aligned with CLUSTALW. Dashes indicate gaps that maximize the alignment, and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with, (*), (:) and (.), respectively. Numbers right to the sequences denote their lengths; those within parenthesis indicate the size of individual domains. Canonical cysteines involved in an intra-chain disulfide bonds within α2 and α3 domains as well as other cysteines mentioned in the text are displayed in inverted type. Conserved residues located at the edges of the binding cleft are highlighted with red shading. Other important residues within the PBD are marked with (▼). Residues involved in salt bridges are denoted with green shading. Residues that contact β2m are highlighted with ($). Residues that contact CD8 (Gao et al., 1997; Kern et al., 1998; Wang et al., 2009) are indicated with yellow shading. A grey bar above the alignment denotes residues involved in tapasin contacts. For each sequence an open box delimits the TM domain.

α3 domain



CP TM CYT domains

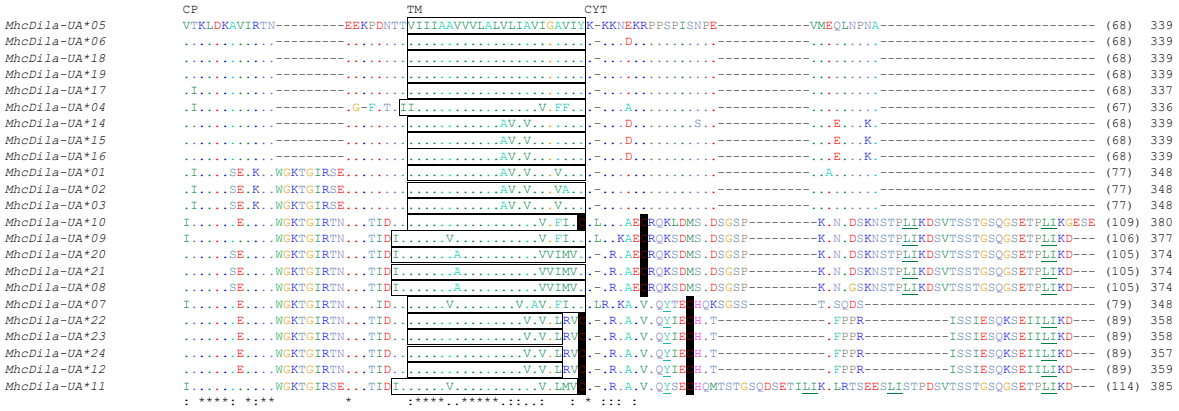


Figure S3 *(continued)*

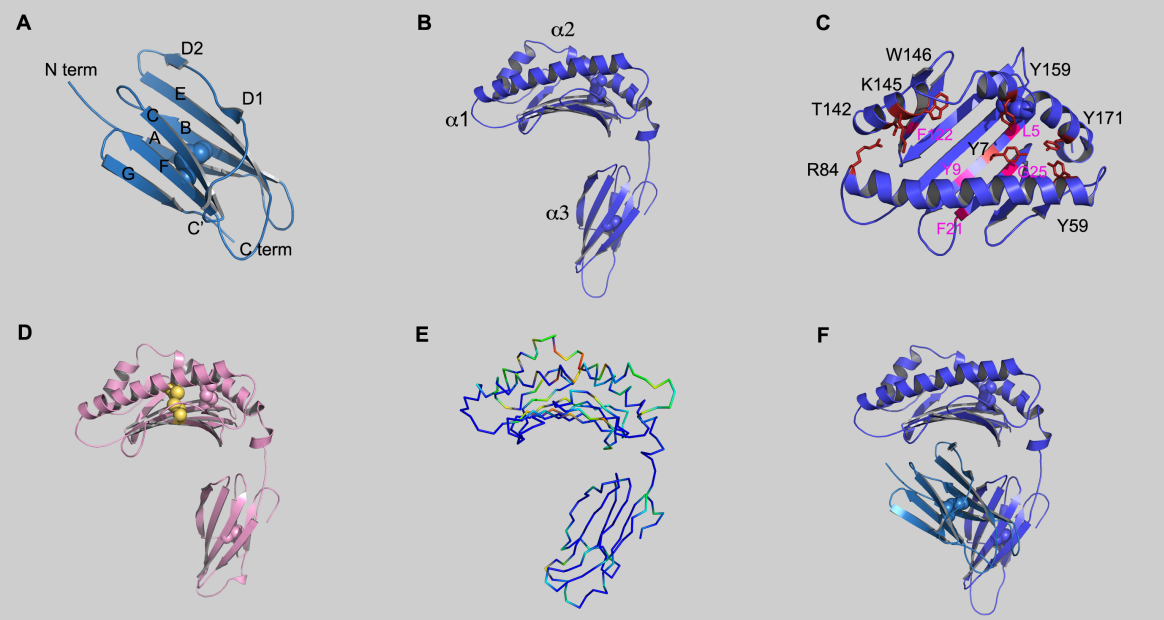


Figure S4. Three-dimensional homology models of sea bass β2m and MHC class I heavy chain extracellular domains. Models generated using the crystal structures of grass carp β2m [PDB: 3GBL] (Chen et al., 2010) and chicken BF2 [PDB: 3BEV] (Koch et al., 2007) as templates. **A) Representation of the Dila-β2m homology model.** Secondary structure elements (SSE), as well as N- and C-termini are labelled. The disulfide bond is represented as spheres. **B) Representation of the predicted Dila-UA*06 model in dark blue.** The three extracellular domains are labelled (α1, α2, α3). Disulfide forming residues are represented as spheres. **C) Top-view of the predicted Dila-UA*06 model,** through a 90º counter-clockwise rotation around the x-axis with respect to the orientation presented in B showing the PBD in detail. The eight anchor residues important for peptide binding are coloured red and labelled and their side-chains are shown. Other conserved residues also contacting the peptide are coloured pink and labelled. **D) Representation of the predicted Dila-UA*12 model in pink.** The cysteine residues represented in yellow could form a putative additional S-S bond. **E) Representation of protein variability among the different sea bass sequences mapped on the Dila-UA*12 homology model** (colour range from dark blue (conserved) to red (variable)). Note that most variable positions are located in the PBD. **F) Representation of the Dila-UA*06/Dila-β2m homology model.** The relative positions of each subunit are based on the chicken BF2 [PDB: 3BEV] (Koch et al., 2007). Heavy and light chains are represented in dark and light blue, respectively. The plasma membrane would be below the α3 domain, while the TCR of a cytotoxic T cell would be on top of the peptide-binding domain (PBD).

CHAPTER 3

Transporters associated with antigen processing (TAP) in sea bass (*Dicentrarchus labrax*, L.): molecular cloning and characterization of TAP1 and TAP2

Rute D. Pinto, Pedro J. B. Pereira and Nuno M. S. dos Santos

Dev Comp Immunol. 2011 35:1173-81

Developmental and Comparative Immunology 35 (2011) 1173–1181

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Short communication

Transporters associated with antigen processing (TAP) in sea bass (*Dicentrarchus labrax*, L.): Molecular cloning and characterization of TAP1 and TAP2

Rute D. Pinto[a,c], Pedro J.B. Pereira[b], Nuno M.S. dos Santos[a,*]

[a] Fish Immunology and Vaccinology Group, IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal
[b] Protein Crystallography Group, IBMC - Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal
[c] ICBAS – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Largo Prof. Abel Salazar 2, 4099-003 Porto, Portugal

ARTICLE INFO

Article history:
Received 2 February 2011
Received in revised form 15 March 2011
Accepted 17 March 2011
Available online 19 April 2011

Keywords:
Dicentrarchus labrax
TAP
Gene cloning
Comparative sequence analysis
Protein structure prediction
Homology modelling

ABSTRACT

The transporters associated with antigen processing (TAP), play an important role in the MHC class I antigen presentation pathway. In this work, sea bass (*Dicentrarchus labrax*) TAP1 and TAP2 genes and transcripts were isolated and characterized. Only the TAP2 gene is structurally similar to its human orthologue. As other TAP molecules, sea bass TAP1 and TAP2 are formed by one N-terminal accessory domain, one core membrane-spanning domain and one canonical C-terminal nucleotide-binding domain. Homology modelling of the sea bass TAP dimer predicts that its quaternary structure is in accordance with that of other ABC transporters. Phylogenetic analysis segregates sea bass TAP1 and TAP2 into each subfamily cluster of transporters, placing them in the fish class and suggesting that the basic structure of these transport-associated proteins is evolutionarily conserved. Furthermore, the present data provides information that will enable more studies on the class I antigen presentation pathway in this important fish species.

1. Introduction

The transporter associated with antigen processing (TAP) is a key component in the MHC class I-dependent antigen presentation pathway [reviewed by Peaper and Cresswell, 2008]. TAP translocates peptides mainly derived from proteasomal cleavage of ubiquitinated proteins in the cytosol to the ER lumen, where these peptides are loaded onto MHC class I molecules (Abele and Tampé, 2009; Michalek et al., 1993). Stable MHC-peptide complexes migrate to the cell surface for presentation to $CD8^+$ cytotoxic T lymphocytes and this recognition event triggers elimination of the presenting cell. TAP is a member of the ABC (ATP-binding cassette) family of transporters, which translocate a wide range of substrates across membranes in an ATP-dependent manner. ABC transporters have a conserved architecture of two transmembrane domains (TMDs) each with six canonical transmembrane helices that together form the translocation pathway and two cytosolic nucleotide-binding domains (NBDs) together bearing two ATP-binding sites. ATP binding induces the NBDs dimerization, with the two ATP molecules sandwiched at the interface and bridging contacts between the two NBDs (Procko et al., 2006). The TAP translocation complex is composed of two half-transporters, TAP1 (ABCB2) and TAP2 (ABCB3), each with one TMD and one NBD (Kelly et al., 1992; Powis et al., 1991; Spies and DeMars, 1991). Additionally, TAP molecules have an N-terminal accessory domain (Procko et al., 2005), also with three to four membrane-spanning regions, that is not required for peptide binding and transport, but is essential for tapasin-binding and assembly of the peptide loading complex (Koch et al., 2004). TAP binds 8–16 amino acid long peptides [reviewed in Abele and Tampé, 2009] but its peptide specificity is restricted to the three N-terminal residues and the C-terminal residue (Uebel et al., 1997). Both TAP subunits were found to be essential and sufficient for ATP-dependent peptide translocation into the ER lumen (Androlewicz et al., 1993; Meyer et al., 1994; Shepherd et al., 1993). TAP genes have been identified and mapped to the MHC in several non-mammalian species, including cartilaginous and bony fish (Grimholt, 1997; Hansen et al., 1999; Ohta et al., 1999a), chicken (Walker et al., 2005) and the amphibian *Xenopus* (Ohta et al., 1999b; Ohta et al., 2003). Here, sea bass TAP1 and TAP2 cDNA and genes have been for the first time identified and characterized at the molecular, structural and phylogenetic levels.

2. Materials and methods

2.1. Fish

Sea bass, *Dicentrarchus labrax*, were kept in a recirculating, ozone-treated salt-water (20–25‰) system at 22 ± 1 °C and fed

* Corresponding author. Tel.: +351 226 074 900; fax: +351 226 099 157.
E-mail addresses: rsp@ibmc.up.pt (R.D. Pinto), ppereira@ibmc.up.pt (P.J.B. Pereira), nsantos@ibmc.up.pt (N.M.S. dos Santos).

0145-305X/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2011.03.024

with commercial pellets twice a day. Fish were sacrificed with a lethal dose of 2-phenoxyethanol (Panreac; >5 mL/10 L).

2.2. cDNA cloning

RNAs were transcribed to cDNAs following the BioScript RNase H⁻ (Bioline) protocol. PCR products were purified from agarose according to QIAquick Gel Extraction Kit (QIAGEN), and cloned following the pGEM®-T Easy Vector Systems instructions (Promega). Automated sequencing was performed using primers detailed in Supplementary Table S1. Complementary DNAs used for 5′ Rapid Amplification of cDNA Ends (5′RACE) experiments were purified using High Pure PCR Product Purification Kit reagents (Roche). For 5′RACE experiments the 5′RACE System from Invitrogen (Version 2.0) was followed and Recombinant Terminal Transferase (Fermentas or Roche) used to dATP tail the purified cDNA.

2.2.1. TAP1

Total RNA, extracted from the spleen of a stimulated fish, was reverse transcribed using primer APv2 (Table S1). The cDNA was first amplified with primers TAP1FW2/TAP1RV1 (Table S1) followed by a semi-nested amplification using primers TAP1FW2/TAP1RV2 (Table S1). The amplification product (~250 bp) obtained was purified, cloned and sequenced.

To obtain the 5′ untranslated region (UTR), three consecutive 5′RACE experiments were performed using cDNA synthesized from previously isolated RNA as described above: (i) the tailed cDNA was first amplified with primers APv/DLTAP1RV2 (Table S1) followed by a second PCR using primers TAP1FW1/DLTAP1RV3 (Table S1); (ii) the tailed cDNA was first amplified with primers APv2/DLTAP1RV8 (Table S1) followed by a second PCR using primers AUAP2/DLTAP1RV9 (Table S1); (iii) the tailed cDNA was first amplified with primers APv/DLTAP1RV11 (Table S1) followed by a second PCR using primers AUAP/DLTAP1RV12 (Table S1). PCR products of ~430 (i), ~750 (ii) and (iii) ~500 bp were consecutively obtained, purified, cloned and sequenced.

To obtain the full sea bass TAP1 cDNA, specific primers were designed at the beginning of the 5′UTR region. The cDNAs from the spleens of two fish were used as templates in parallel PCR reactions. Primers DLTAP1FW1/APv2 (Table S1) were used in a first amplification PCR (spleen cDNAs from two differently stimulated fish) with additional 0.15 U of *Pfu* DNA polymerase (Promega) followed by a nested PCR using primers DLTAP1FW2/AUAP or AUAP2 (Table S1). The obtained ~3000 bp product was purified, cloned and 2 independent clones from each template were sequenced (Table S1).

2.2.2. TAP2

Total RNA extracted from the head kidneys of three stimulated fish was reverse transcribed using primer APv2 (Table S1). The cDNA was amplified with primers TAP2FW1/TAP2RV1 (Table S1) followed by a semi-nested PCR using primers TAP2FW1/TAP2RV2 (Table S1). A third amplification was performed with primers TAP2FW2/TAP2RV2 (Table S1). The obtained product of ~970 bp was purified, cloned, and sequenced (Table S1).

This sequence was used to design three specific reverse primers in order to follow a 5′RACE strategy: DLTAP2RV1, DLTAP2RV2, DLTAP2RV3 (Table S1). The previously isolated RNA was used to synthesize first-strand cDNA with primer DLTAP2RV1. The cDNA was purified, tailed, and first amplified with the primers APv/DLTAP2RV2. A second PCR was performed with primers AUAP/DLTAP2RV3 (Table S1) and a product of ~800 bp was obtained, purified, cloned and sequenced.

To obtain the full sea bass TAP2 cDNA, specific primers (Table S1) were designed at the beginning of the 5′UTR region. DLTAP2FW1 and AUAP2 were used in a first amplification using unstimulated head kidney cDNA, followed by a semi-nested PCR with primers DLTAP2FW2/AUAP2. The obtained ~2500 bp PCR product was purified, cloned and 3 independent clones from each template were sequenced (Table S1).

2.3. Genomic DNA cloning

Genomic DNA (gDNA) was isolated from sea bass erythrocytes from a single fish, as described by Stet et al. (1993). To obtain the full TAP1 and TAP2 genes, gDNA was amplified by PCR with DLTAP1FW1/DLTAP1RV14 and DLTAP2FW1/DLTAP2RV6 (Table S1), respectively. The PCR products were purified, cloned and one clone of each molecule sequenced as before (Table S1).

2.4. Sequence analysis

Full nucleotide and protein sequences from sea bass TAP molecules were compared to several TAP1 and TAP2 sequences currently available in the GenBank database [http://www.ncbi.nlm.nih.gov/genbank/]. The multiple sequence alignment was made using CLUSTALW (Higgins, 1994) and formatted with Bioedit (Hall, 1998). The transmembrane/nucleotide-binding domain boundaries were based on InterPro Scan predictions (Zdobnov and Apweiler, 2001). The molecular weights of the polypeptide chains were calculated with the Expasy compute pI/MW tool [http://www.expasy.org/tools/pi_tool.html]. Possible N-glycosylation sites were predicted with Expasy post-translational modification tool [http://www.cbs.dtu.dk/services/NetNGlyc/]. The neighbour-joining phylogenetic tree was constructed with MEGA v3.1 (Kumar et al., 2004), using p-distance parameter and complete deletion of gaps, and tested for reliability using 2000 bootstrap replications. The percentages of similarity and identity were calculated with MatGAT (Campanella et al., 2003) using default parameters.

2.5. Southern blotting

Genomic DNA was digested with different restriction enzymes (EcoRI, NcoI and NdeI for TAP1; EcoRI, PstI, and NdeI for TAP2 all zero cutters within the probe), separated in a 0.8% agarose gel and subjected to Southern blotting (Ausubel et al., 1999). A portion of the sea bass TAP1 and TAP2 cDNA (partial exon I) was amplified with primers DLTAP1FW2/DLTAP1RV12 and DLTAP2FW2/DLTAP2RV7 (Table S1), labelled and used as probes.

Preparation of the labelled probes, hybridization and post-hybridization stringency washes were performed accordingly to Gene Images™ AlhPhos Direct™ Labelling and Detection System (Amersham Biosciences) kit. For signal generation and detection the Chemiluminescent Signal Generation and Detection with CDP-*Star*™ protocol from the same kit was followed.

2.6. 3D modelling

Two suitable structural templates for Dila-TAP1 and Dila-TAP2, Sav1866 (PDB: 2hyd) and mouse P-glycoprotein (PDB: 3g61), were identified by a BLAST search as implemented in the SWISS-MODEL Protein Modelling Server (Arnold et al., 2006). A multiple sequence alignment was used for homology modelling with SWISS-MODEL in alignment mode (Arnold et al., 2006) revealing identities of approximately 30% between target and template molecules with the gaps introduced essentially outside the secondary structure elements (SSE) and SSE were identified with PROMOTIF (Hutchinson and Thornton, 1996). The calculated global energies were of −17,356.039 kJ/mol (TAP1-2hyd), −12,068.762 kJ/mol (TAP2-2hyd), −10,095.534 kJ/mol (TAP1-3g61_C-terminal) and

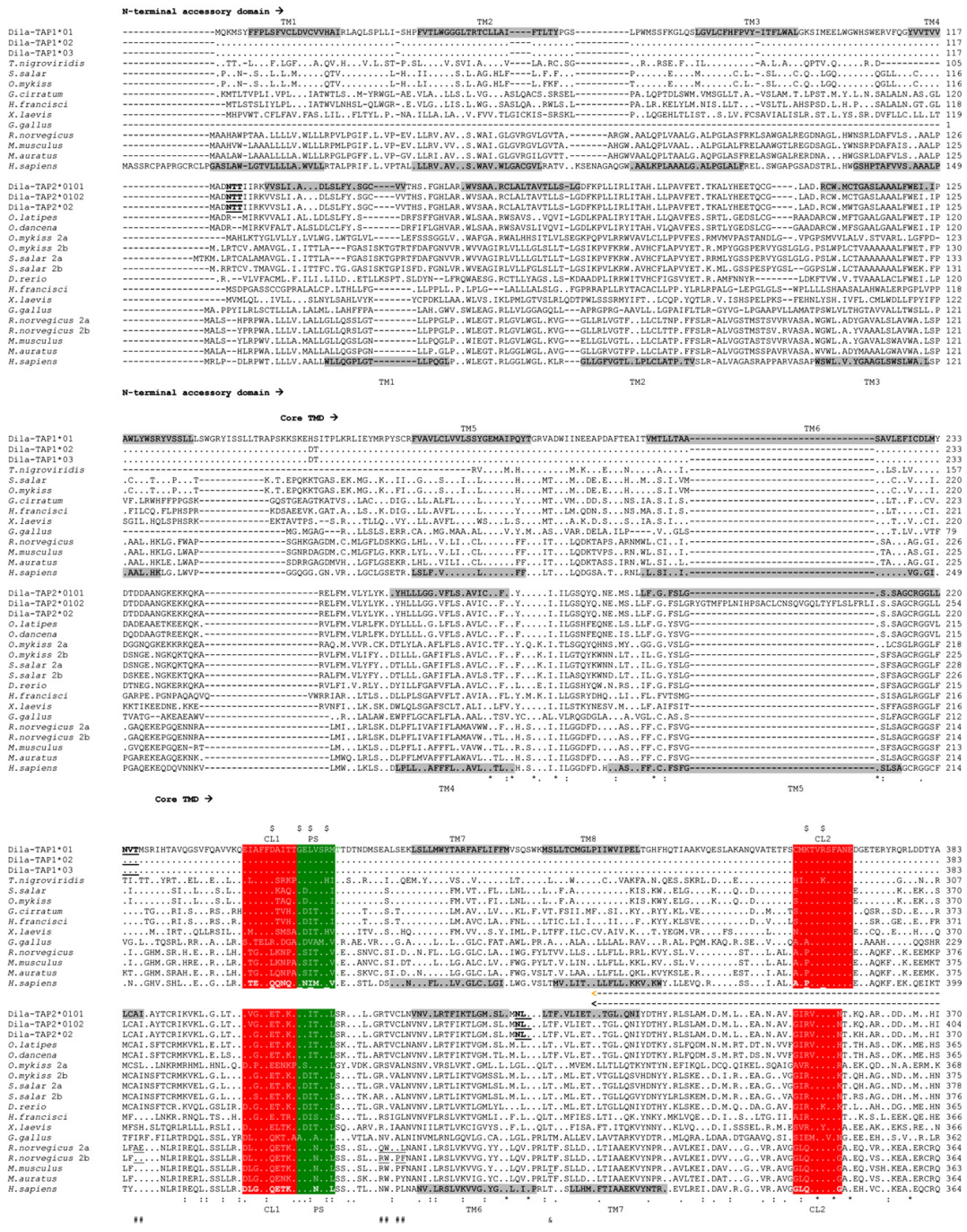


Fig. 1. Alignment of TAP1 and TAP2 amino acid sequences. The TAP1/TAP2 protein sequences were aligned using CLUSTALW. As for other TAPs, no putative N-terminal leader sequence was predicted for Dila-TAP1 and -TAP2, nor was any ER retention motif identified in any of the sea bass subunits. Membrane topology and structural features are based on human TAP1 and on a 3D homology model of the human core TAP complex, respectively (Oancea et al., 2009; Schrodt et al., 2006). The proposed N-terminal accessory domain, core transmembrane domain, nucleotide-binding domain are indicated above and below the alignment; membrane-spanning helices are shaded in grey; and several other features including CL1 and CL2 (shaded in red), peptide sensor (shaded in green), A-loop, Q-loop, X-loop, C-loop, D-loop, H-loop and Walker A and B motifs (all shaded in blue) are denoted above and bellow the alignment, for TAP1 and TAP2. Peptide binding regions 1 (PBR1) and 2 (PBR2) are indicated between TAP1 and TAP2 alignments (<---> for TAP1 <- - -> for TAP2). Putative N-glycosylation sites are in bold type and underlined, for both Dila-TAP1 and -TAP2 sequences. Location of particular residues mentioned in the text is denoted with symbols (# – rat, $ – human, and – mouse) either above or below the alignment for TAP1 or TAP2 respectively, and the residues are

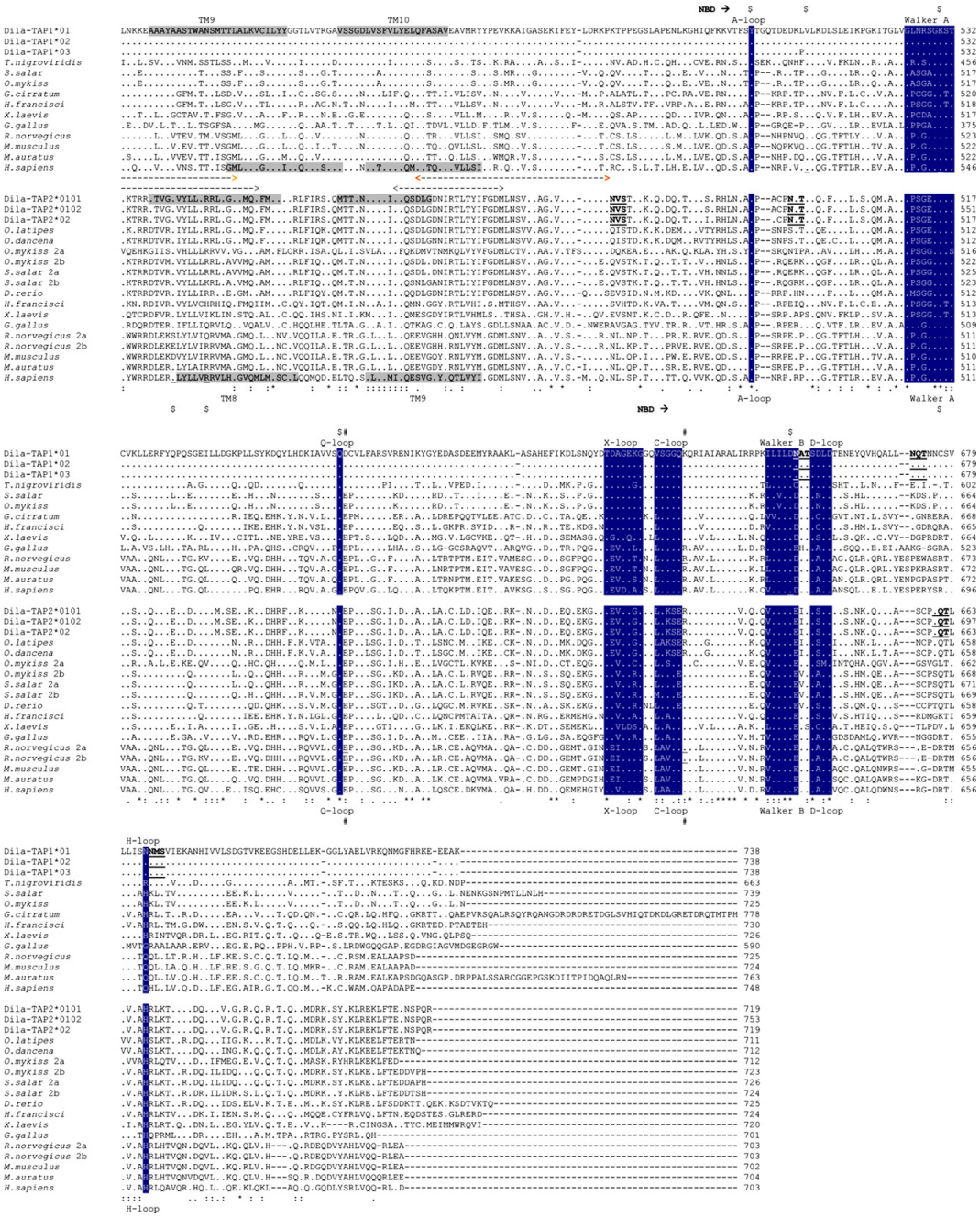


Fig. 1. (*Continued*).

underlined. Dashes indicate gaps introduced to optimise similarity between sequences. Dots indicate identity to Dila-TAP1*01 sequence. Asterisks denote residues identical in all sequences. Double dots and dots indicate chemical similarity between amino acids, i.e., conserved and semi-conserved substitutions, respectively.

−1176.007 kJ/mol (TAP2-3g61_N-terminal). Two dimeric models were generated by superposing the TAP1 and TAP2 homology models with each template chains (2hyd) or with N- and C-terminal regions of the unique template's chain (3g61), and energies minimized using PHENIX (Adams et al., 2010), with resulting QMEAN scores of 0.801 and 0.667, respectively. Model quality was assessed with WHATCHECK (Hooft et al., 1996) and PROCHECK (Laskowski et al., 1993) as implemented in the Swiss Model structure assessment tool. Model pictures were prepared with PyMOL [http://pymol.org/]. Additionally, for comparison, structures of the isolated NBD domains of Dila-TAP1 and -TAP2 were also predicted using the crystal structure of the dimeric form of rat TAP1 NBD [PDB: 2ixf]. The calculated global energies were −10,590.928 kJ/mol (TAP1-2ixf) and −11,170.671 kJ/mol (TAP2-2ixf).

3. Results and discussion

3.1. cDNA sequence

Two different TAP1 cDNAs were obtained from sea bass (Dila-TAP1) (Fig. S1) and named Dila-TAP1*01 (3188 bp) [GenBank: HQ328078] and Dila-TAP1*02 (3187 bp) [GenBank: HQ328079]. BLAST analysis confirmed higher similarity with fish TAP1. Both TAP1 sequences have an open reading frame (ORF) of 2217 bp, a 5′UTR of 136 bp, and a 3′UTR of 834/5 bp excluding the polyadenylation tail. Within the 3′UTR there are three potential polyadenylation signals (one non-canonical) and two mRNA instability motifs (Shaw and Kamen, 1986) (Fig. S1.1A).

Two different TAP2 cDNAs were obtained from sea bass (Dila-TAP2) (Fig. S1) and named Dila-TAP2*0101 (2563 bp) [GenBank: HQ328081] and Dila-TAP2*0102 (2605 bp) [GenBank: HQ328082]. BLAST analysis confirmed higher similarity with fish TAP2. The here obtained sea bass TAP2 sequences have an ORF of 2160 bp (TAP2*0101) or 2262 bp (TAP2*0102), a 5′UTR of 24 bp, and a 3′UTR of 397 bp excluding the polyadenylation tail. The larger transcript is a splice variant where intron 2 is incorporated into the coding region (see below). Relevance of this sea bass TAP2 isoform remains to be elucidated, although two splicing forms of human TAP2 have been described and associated to differential peptide selectivity (Yan et al., 1999). Furthermore, in humans there is evidence for a recombinatorial hot spot in intron 2 of the TAP2 locus (Cullen et al., 1995). Within 3′UTR there are three potential polyadenylation signals and one mRNA instability motif (Fig. S1.1B).

3.2. Genomic DNA sequence, structure and copy number

The Dila-TAP1 gene (7864 bp) was obtained from a third fish and named Dila-TAP1*03 [GenBank: HQ328080] (Fig. S1.2.A). Six single nucleotide differences were found between this sequence and those just described above for TAP1 cDNAs (Fig. S1.2.A). The Dila-TAP1 gene contains 13 exons and all intron donor/acceptor sites flanking the exons match the consensus motifs. Dila-TAP1 13-exon structure is different from that of the 11-exon human TAP1 (Beck et al., 1992) (Fig S2A). The sea bass molecule has two additional introns, one within the 5′UTR and the other within human exon I, but both the location and phase of the remaining introns are conserved between human and sea bass. Chicken TAP1 gene organization is identical to the human one (Walker et al., 2005), and so sea bass TAP1 is, for now, unique in its structure.

Dila-TAP2 gene (5674 bp), obtained from another individual was named Dila-TAP2*02 [GenBank: HQ328083] (Fig. S1.2.B). It has an 11-exon structure similar to that of the human gene (Beck et al., 1992) and distinct from those of chicken (Walker et al., 2005) and Atlantic salmon (Grimholt, 1997), which were described as 9 and 10 exons genes, respectively. All intron donor/acceptor sites follow the gt/ag rule (Fig. S1.2B), and both the location and phase of all the introns are conserved between human and sea bass TAP2 molecules (Fig. S2B). This is also true for Atlantic salmon TAP2, except for intron 10, which is absent in this species (Grimholt, 1997).

Regarding the number of copies of genes in the genome, Southern blot results suggest that both Dila-TAP1 and Dila-TAP2 are encoded by single genes (Fig. S3A). Although RFLPs (restriction fragment length polymorphisms) were observed for Dila-TAP2 between the two analysed fish (Fig. S3B), they are most likely due to allelic polymorphism. Of note, two types of TAP2 (TAP2A and TAP2B) have been reported in Atlantic salmon (Grimholt, 1997), and in rainbow trout (Hansen et al., 1999), with one or both of them being polymorphic.

3.3. Primary structure analysis

In order to compare TAP proteins from sea bass and other vertebrates (Table S2), a multiple sequence alignment was made using CLUSTALW (Higgins, 1994). As observed for other organisms, Dila-TAP1 protein (83 kDa) is slightly larger than Dila-TAP2*01 (80 kDa), but about the same molecular mass as Dila-TAP2*02 (84 kDa). Variation in size can be observed in both molecules across species, perhaps with more heterogeneity within TAP1 proteins (Fig. 1). Dila-TAP2*02 is considerably longer than all other TAP2 molecules described so far. Putative N-glycosylation sites were identified in TAP1 (four) and TAP2 (five) as reported for Atlantic salmon (four for TAP1 and one for TAP2) (Grimholt, 1997) and human (four for TAP1).

A number of transmembrane α-helices were predicted within both the N-terminal accessory domain and the TMD of sea bass TAPs (Swiss Model Domain Annotation Tool), closely resembling those predicted for the human molecules (10 for Dila-TAP1 and 9 for Dila TAP2). In addition, Dila-TAP1 and -TAP2 have a conserved ATP binding region, consistent with other transporters.

The predicted N-terminal tapasin-binding domain (or N domain) of TAP1 and TAP2 consists of 4 and 3 TM regions, respectively. This is a poorly conserved region (no invariant residues) in the multiple sequence alignment, to which have been attributed several biochemical interactions including heterodimer formation (Vos et al., 1999) and tapasin recruitment (Garbi et al., 2003; Koch et al., 2004). In comparison to TAP2, TAP1 displays a number of insertions at the final section of the N-domain, which are more notorious in the sea bass molecules than in other species.

The core TMD includes the remaining membrane-spanning regions, six for each sea bass half-transporter. Within this domain there are 18 invariant residues between all TAP sequences (Fig. 1). Clusters of amino acids that can alter peptide specificity have been identified in rat TAP2 core TMD (Joly and Butcher, 1998), restoring permissive transport in the restricted TAP2-2b chain (Fig. 1). These regions are variable in TAP2 across species, but tend to be conserved in TAP1. A mutation in mouse TAP2 ($T^{293}P$) is responsible for a reduction in TAP2 protein levels in the recent Jasmine strain (Theodoratos et al., 2010). This residue is conserved in TAP2 across species pointing to a conserved role in TAP stabilization. The TMD of human TAP molecules includes two peptide-binding regions (PBR) (Nijenhuis and Hammerling, 1996; Nijenhuis et al., 1996), which also show conservation across species. Considering TAP1 and TAP2 separately, there is 45% and 37% identity in PBR1 and 65% and 50% in PBR2 between human and sea bass molecules. In human TAP2, a single amino acid substitution ($A^{374}D$), changes the preference for the transported peptide (Armandola et al., 1996). This residue is conserved in all species, except rodents (serine or glutamate in rat; aspartate in mouse). TAP2 R^{380} (Momburg et al., 1996) also influences transport and is present in most species including sea bass, but not in rat and chicken (glutamine), amphibian (isoleucine) and

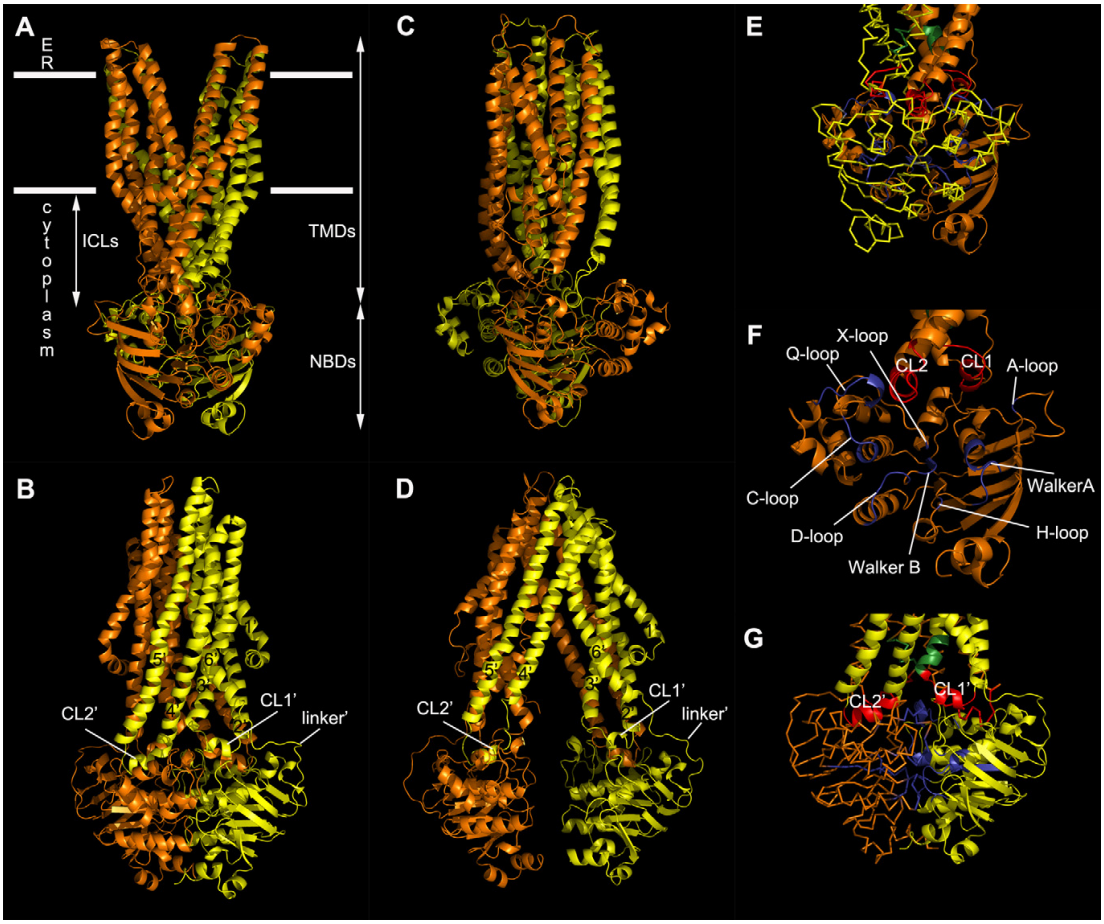


Fig. 2. 3D homology models of the accessory domain-less Dila-TAP1/-TAP2 heterodimer. Models generated based on the crystal structures of Sav1866 [PDB: 2hyd] (A, B, E, F and G) and mouse ABCB1 [PDB: 3g61] (C and D) in close outward-facing (ATP-bound) and open inward-facing (ATP-free) conformations, respectively. Dila-TAP1 is in orange and Dila-TAP2 in yellow. For both molecules, coupling loops are colored red, the peptide sensor green and conserved motifs of the NBD blue (E, F and G). (A) Horizontal lines indicate the approximate location of the ER membrane, ER lumen on top and cytoplasm below. Vertical arrows denote the transmembrane domains (TMDs), nucleotide-binding domains (NBDs) and the intracytoplasmic loops (ICLs). This orientation highlights the outward-facing conformation of the molecule, with the translocation cavity open towards the ER lumen. (B) 90° clockwise rotation around the *y*-axis with respect to the orientation presented in A. This orientation highlights two wings of the complex composed of helices 1, 2, 3 and 6 from one chain and helices 4′ and 5′ from the other (and vice versa) and the NBDs interface in close contact. For TAP2, TM helices are numbered 1′–6′; coupling helices (CL1′ and CL2′) that assist conformational changes, and also the loop (linker') that connects the two domains are indicated. (C) Heterodimer based on ABCB1. (D) 90° clockwise rotation around the *y*-axis with respect to the orientation presented in C. Highlighted the inward-facing conformation of the molecule, with the translocation cavity open towards the cytoplasm. NBDs are not in contact with each other. (E) Close-up representation of TAP1/TAP2 dimer rotated 180° around the *y*-axis as depicted in A. TAP2 molecule is shown as Cα trace. (F) Close-up representation of TAP1 NBD interface rotated 180° around the *y*-axis as depicted in A. The conserved motifs of the NBD are labelled. (G) Close-up representation of the NBDs interface as depicted in B. TAP1 molecule is shown as Cα trace.

shark (histidine). Both residues are located in the putative peptide binding/release cavity. Each half-transporter has two coupling helices or loops, CH1/CL1 and CH2/CL2, and a peptide sensor site has been identified close to the CL1 (Herget et al., 2007). Mutations on highly conserved residues of CH1/CL1 (except $Q^{277}C$) and of sensor region ($G^{282}C$, $I^{284}C$, $R^{287}C$) interfered with and drastically affected transport, respectively. In CH2/CL2, only $R^{378}C$ affected transport and binding and $P^{375}C$ influenced transport but not binding (Oancea et al., 2009).

Globally, the NBDs are the most conserved with the eight consensus motifs of ABC transporters required for ATP binding and hydrolysis [reviewed by Kos and Ford, 2009], present in Dila-TAP1 and -TAP2, confirming the existence of two ATPase sites in the sea bass complex. The tyrosine in the A-loop that helps positioning ATP in the catalytic site [reviewed in Kos and Ford, 2009] is 100% conserved in both molecules across species. The Walker A motif is conserved across all TAP2 and in Dila-TAP1 displays a single residue variation (R instead of G at position 4) observed also in *Tetraodon* and *Xenopus*. Three of the residues, including hTAP2 K^{509}/hTAP1 K^{544}, a possible ATP interaction site (Ohta et al., 1999b) are absolutely invariant. The glutamine in the Q-loop thought to sense the γ-phosphate and to communicate with the TMDs (Seeger and van Veen, 2009) is totally conserved. The glutamate in the X-loop that also plays an important role in the translocation event, but not in substrate binding (Oancea et al., 2009), is present in all sequences except for *Xenopus* TAP1/2 (glutamine/aspartate). The D-loop (SALD) aspartate interacts with the Walker A motif while alanine contacts the H-loop in trans-NBD (Seeger and van Veen,

2009). The aspartate is totally invariant across species, but the alanine is not (aspartate/serine in sea bass TAP1/TAP2, respectively), which could be explained by the nature of the interaction (alanine main-chain to histidine side chain). The C-loop, hallmark of ABC proteins, is conserved in TAP1 across species (only conservative replacement in teleosts), displaying variation among TAP2. The Walker B motif (ϕϕϕϕDE, ϕ any aliphatic residue) (Walker et al., 1982) contacting both γ-phosphate and magnesium ion through the aspartate (hTAP2 D^{631}/hTAP1 D^{667}) (Hopfner et al., 2000; Kos and Ford, 2009) shows variation in Dila-TAP1 while in Dila-TAP2, this motif is 100% conserved comparing to hTAPs. Finally, the histidine (H^{661}, hTAP2) in the H-loop, present in all TAP2, was considered essential for transport in other ABC transporters (Bliss et al., 1996; Shyamala et al., 1991). Variation is observed in TAP1 across species: a glutamine (Q^{701}, hTAP1) present in all mammalian TAP1 was shown to interact with the γ-phosphate in human TAP1 (Gaudet and Wiley, 2001); in sea bass an asparagine replaces the very similar glutamine observed in mammals; in pufferfish an arginine occupies this position, being more similar to the histidine observed in other fish; surprisingly, chicken displays a glycine at this position.

Noteworthy, in the context of the sea bass NBDs dimer, almost all deviations map to the same ATPase site: [Walker A + Walker B + H-loop]^{TAP1} + [D-loop + Signature motif]^{TAP2}. This is in accordance with previous reports [reviewed in Procko et al., 2009], where one ATPase site is formed by consensus residues, while the second contains non-consensus substitutions in one or more motifs (degenerate site), a common property among eukaryotic ABC transporters (Procko and Gaudet, 2009).

3.4. 3D modelling

The Dila-TAP1 and -TAP2 structures (except their accessory domains) were predicted by comparative modelling using the Sav1866 crystal structure [PDB: 2hyd] and the mouse P-glycoprotein crystal structure [PDB: 3g61] as templates, respectively (Figs. 2 and S4). The accessory domain-less homology models of Dila-TAP1 and -TAP2 are compatible with the presence of six TM α-helices in the TMD, followed by a canonical ABC-transporter NBD, as described for other ABC half-transporters (Fig. 2A–D), with spatial conservation of important structural amino acids (Figs. 2E–G and S4). The models of the NBDs, including those calculated using the crystal structure of the rTAP1 NBD in dimeric form [PDB: 2ixf], display significant homology with equivalent domains from homologous molecules, with a RecA-like sub-domain and a helical sub-domain in both Dila-TAP1 and -TAP2 (Fig. S4). Finally, the quaternary heterodimeric sea bass TAP models were generated by superposing the homology models of the individual subunits onto two different templates, Sav1866 and ABCB1, in outward/closed (nucleotide-bound) and inward/open (nucleotide-free) conformations, respectively. In the quaternary models, the patches of exposed hydrophobic residues are compatible with the position occupied by the transmembrane regions in homologous molecules, and a number of conserved (and presumably functionally important) residues map to functionally relevant areas of the heterodimer (detailed in supplementary section).

3.5. Phylogenetic analysis

TAP1 and TAP2 protein sequences were phylogenetically compared to each other and also to protein sequences of TAPL or ABCB9 (Table S2), another member of the eukaryote ABC-transporter family (Fig. 3). The neighbour-joining tree clearly separates the three types of amino acid sequences analysed, placing the Dila-TAP1 and the Dila-TAP2 molecules within each respective sub-family. According to the method used, the sea bass TAP1 clusters with other

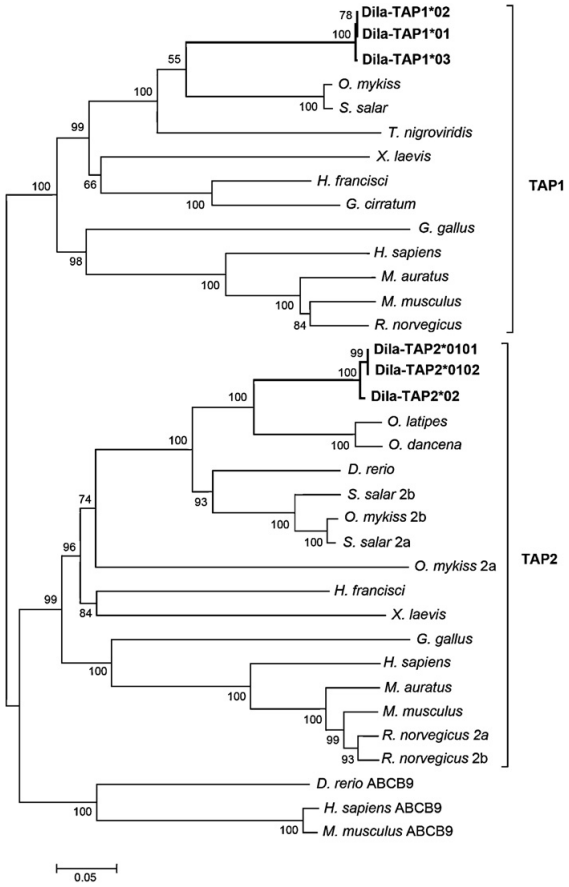


Fig. 3. Phylogenetic tree of TAP1, TAP2 and TAPL. An amino acid alignment (CLUSTALW) of the proteins was used to generate the unrooted neighbour-joining tree with MEGA v3.1. Node values represent percentages of bootstrap analysis from 2000 replications and complete deletion of gaps. The TAP1 and TAP2 sequences used are those indicated in Fig. 1, while for TAPL/ABCB9 are: [GenBank: Q9NP78] (human), [GenBank: Q9JJ59] (mouse) and [GenBank: CAM46975] (zebrafish).

bony fish molecules, being more closely related to the salmonid ones. In respect to sea bass TAP2, this molecule is also placed in the teleost group, and is more closely related to those from *O. latipes* and *O. dancena*. Additionally, pair-wise alignments were performed with MatGAT (Campanella et al., 2003) in order to calculate similarity and identity of TAP amino acids between species. The results confirm those previously obtained by the neighbour-joining method, being sea bass TAP1 apparently more closely related to *S. salar* and to *O. mykiss*, and sea bass TAP2 molecules closer to those of *O. latipes* and *O. dancena* (Table S2). The percentages of similarity and identity of Dila-TAP1 and -TAP2 to each other (58% and 36%, respectively) are lower than those observed between each of them and the respective subfamily cluster, indicating that these are indeed distinct molecules within this species.

4. Conclusions

In the present study, the sea bass TAP1 and TAP2 transcripts were successfully cloned for the first time. Each consists of one N-terminal accessory domain, one TMD (both with multiple transmembrane helices) and one NBD located in the cytoplasm. They

are homologous to the human proteins and several structural features resemble those of other TAPs, namely two coupling loops in the transmission interface of each half-transporter and two ATPase sites (conserved and degenerate) in the NBDs interface, suggesting a similar function. Importantly, homology modelling suggests a heterodimeric ABC transporter-like quaternary architecture for sea bass TAPs. Noteworthy, a TAP2 splice variant has been identified, which includes intron 2. Gene structure has been determined for both sea bass TAP molecules, with Dila-TAP1 displaying a unique organization, and Dila-TAP2 organized as its human homologue; both molecules are encoded by single-copy genes. Finally, Dila-TAP1 and -TAP2 are phylogenetically segregated in each sub-family, being closer to other bony fish molecules. Hence, the present identification of sea bass TAP genes provides information that will allow further studies on class I antigen presentation in this important fish species.

Acknowledgements

This work was supported by funding from FCT - Fundação para a Ciência e a Tecnologia, Portugal (POCTI/CVT/44925/2006). Rute D. Pinto acknowledges FCT PhD fellowship BD/42327/2007, financed by POPH-QREN and co-funded by FSE and MCTES. The funders had no role in study design, data collection and analysis, decision to publish, or publication of the manuscript. No additional external funding was received for this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.dci.2011.03.024.

References

Abele, R., Tampé, R., 2009. Peptide trafficking and translocation across membranes in cellular signaling and self-defense strategies. Curr. Opin. Cell Biol. 21, 508–515.
Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D: Biol. Crystallogr. 66, 213–221.
Androlewicz, M.J., Anderson, K.S., Cresswell, P., 1993. Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. Proc. Natl. Acad. Sci. U.S.A. 90, 9130–9134.
Armandola, E.A., Momburg, F., Nijenhuis, M., Bulbuc, N., Fruh, K., Hammerling, G.J., 1996. A point mutation in the human transporter associated with antigen processing (TAP2) alters the peptide transport specificity. Eur. J. Immunol. 26, 1748–1755.
Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195–201.
Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., 1999. In: Chanda, V. (Ed.), Current Protocols in Molecular Biology. John Wiley and Sons, Inc., New York.
Beck, S., Kelly, A., Radley, E., Khurshid, F., Alderton, R.P., Trowsdale, J., 1992. DNA sequence analysis of 66 kb of the human MHC class II region encoding a cluster of genes for antigen processing. J. Mol. Biol. 228, 433–441.
Bliss, J.M., Garon, C.F., Silver, R.P., 1996. Polysialic acid export in *Escherichia coli* K1: the role of KpsT, the ATP-binding component of an ABC transporter, in chain translocation. Glycobiology 6, 445–452.
Campanella, J.J., Bitincka, L., Smalley, J., 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinform. 4, 29.
Cullen, M., Erlich, H., Klitz, W., Carrington, M., 1995. Molecular mapping of a recombination hotspot located in the second intron of the human TAP2 locus. Am. J. Hum. Genet. 56, 1350–1358.
Garbi, N., Tiwari, N., Momburg, F., Hammerling, G.J., 2003. A major role for tapasin as a stabilizer of the TAP peptide transporter and consequences for MHC class I expression. Eur. J. Immunol. 33, 264–273.
Gaudet, R., Wiley, D.C., 2001. Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. EMBO J. 20, 4964–4972.
Grimholt, U., 1997. Transport-associated proteins in Atlantic salmon (*Salmo salar*). Immunogenetics 46, 213–221.
Hall, T., 1998. BioEdit. In: Biological Sequence Alignment Editor for Windows. North Carolina State University, NC, USA.
Hansen, J.D., Strassburger, P., Thorgaard, G.H., Young, W.P., Du Pasquier, L., 1999. Expression, linkage, and polymorphism of MHC-related genes in rainbow trout, *Oncorhynchus mykiss*. J. Immunol. 163, 774–786.
Herget, M., Oancea, G., Schrodt, S., Karas, M., Tampe, R., Abele, R., 2007. Mechanism of substrate sensing and signal transmission within an ABC transporter: use of a Trojan horse strategy. J. Biol. Chem. 282, 3871–3880.
Higgins, D.G., 1994. CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol. Biol. 25, 307–318.
Hooft, R.W., Vriend, G., Sander, C., Abola, E.E., 1996. Errors in protein structures. Nature 381, 272.
Hopfner, K.P., Karcher, A., Shin, D.S., Craig, L., Arthur, L.M., Carney, J.P., Tainer, J.A., 2000. Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. Cell 101, 789–800.
Hutchinson, E.G., Thornton, J.M., 1996. PROMOTIF—a program to identify and analyze structural motifs in proteins. Protein Sci. 5, 212–220.
Joly, E., Butcher, G.W., 1998. Why are there two rat TAPs? Immunol. Today 19, 580–585.
Kelly, A., Powis, S.H., Kerr, L.A., Mockridge, I., Elliott, T., Bastin, J., Uchanska-Ziegler, B., Ziegler, A., Trowsdale, J., Townsend, A., 1992. Assembly and function of the two ABC transporter proteins encoded in the human major histocompatibility complex. Nature 355, 641–644.
Koch, J., Guntrum, R., Heintke, S., Kyritsis, C., Tampe, R., 2004. Functional dissection of the transmembrane domains of the transporter associated with antigen processing (TAP). J. Biol. Chem. 279, 10142–10147.
Kos, V., Ford, R.C., 2009. The ATP-binding cassette family: a structural perspective. Cell. Mol. Life Sci. 66, 3111–3126.
Kumar, S., Tamura, K., Nei, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.
Laskowski, R.A.M.M.W., Moss, D., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283–291.
Meyer, T.H., van Endert, P.M., Uebel, S., Ehring, B., Tampe, R., 1994. Functional expression and purification of the ABC transporter complex associated with antigen processing (TAP) in insect cells. FEBS Lett. 351, 443–447.
Michalek, M.T., Grant, E.P., Gramm, C., Goldberg, A.L., Rock, K.L., 1993. A role for the ubiquitin-dependent proteolytic pathway in MHC class I-restricted antigen presentation. Nature 363, 552–554.
Momburg, F., Armandola, E.A., Post, M., Hammerling, G.J., 1996. Residues in TAP2 peptide transporters controlling substrate specificity. J. Immunol. 156, 1756–1763.
Nijenhuis, M., Hammerling, G.J., 1996. Multiple regions of the transporter associated with antigen processing (TAP) contribute to its peptide binding site. J. Immunol. 157, 5467–5477.
Nijenhuis, M., Schmitt, S., Armandola, E.A., Obst, R., Brunner, J., Hammerling, G.J., 1996. Identification of a contact region for peptide on the TAP1 chain of the transporter associated with antigen processing. J. Immunol. 156, 2186–2195.
Oancea, G., O'Mara, M.L., Bennett, W.F., Tieleman, D.P., Abele, R., Tampe, R., 2009. Structural arrangement of the transmission interface in the antigen ABC transport complex TAP. Proc. Natl. Acad. Sci. U.S.A. 106, 5551–5556.
Ohta, Y., Haliniewski, D.E., Hansen, J., Flajnik, M.F., 1999a. Isolation of transporter associated with antigen processing genes, TAP1 and TAP2, from the horned shark Heterodontus francisci. Immunogenetics 49, 981–986.
Ohta, Y., Powis, S.J., Coadwell, W.J., Haliniewski, D.E., Liu, Y., Li, H., Flajnik, M.F., 1999b. Identification and genetic mapping of Xenopus TAP2 genes. Immunogenetics 49, 171–182.
Ohta, Y., Powis, S.J., Lohr, R.L., Nonaka, M., Pasquier, L.D., Flajnik, M.F., 2003. Two highly divergent ancient allelic lineages of the transporter associated with antigen processing (TAP) gene in Xenopus: further evidence for co-evolution among MHC class I region genes. Eur. J. Immunol. 33, 3017–3027.
Peaper, D.R., Cresswell, P., 2008. Regulation of MHC class I assembly and peptide binding. Annu. Rev. Cell Dev. Biol. 24, 343–368.
Powis, S.J., Townsend, A.R., Deverson, E.V., Bastin, J., Butcher, G.W., Howard, J.C., 1991. Restoration of antigen presentation to the mutant cell line RMA-S by an MHC-linked transporter. Nature 354, 528–531.
Procko, E., Ferrin-O'Connell, I., Ng, S.L., Gaudet, R., 2006. Distinct structural and functional properties of the ATPase sites in an asymmetric ABC transporter. Mol. Cell 24, 51–62.
Procko, E., Gaudet, R., 2009. Antigen processing and presentation: TAPping into ABC transporters. Curr. Opin. Immunol. 21, 84–91.
Procko, E., O'Mara, M.L., Bennett, W.F., Tieleman, D.P., Gaudet, R., 2009. The mechanism of ABC transporters: general lessons from structural and functional studies of an antigenic peptide transporter. FASEB J. 23, 1287–1302.
Procko, E., Raghuraman, G., Wiley, D.C., Raghavan, M., Gaudet, R., 2005. Identification of domain boundaries within the N-termini of TAP1 and TAP2 and their importance in tapasin binding and tapasin-mediated increase in peptide loading of MHC class I. Immunol. Cell Biol. 83, 475–482.
Schrodt, S., Koch, J., Tampe, R., 2006. Membrane topology of the transporter associated with antigen processing (TAP1) within an assembled functional peptide-loading complex. J. Biol. Chem. 281, 6455–6462.
Seeger, M.A., van Veen, H.W., 2009. Molecular basis of multidrug transport by ABC transporters. Biochim. Biophys. Acta 1794, 725–737.

Shaw, G., Kamen, R., 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 46, 659–667.
Shepherd, J.C., Schumacher, T.N., Ashton-Rickardt, P.G., Imaeda, S., Ploegh, H.L., Janeway Jr., C.A., Tonegawa, S., 1993. TAP1-dependent peptide translocation in vitro is ATP dependent and peptide selective. Cell 74, 577–584.
Shyamala, V., Baichwal, V., Beall, E., Ames, G.F., 1991. Structure-function analysis of the histidine permease and comparison with cystic fibrosis mutations. J. Biol. Chem. 266, 18714–18719.
Spies, T., DeMars, R., 1991. Restored expression of major histocompatibility class I molecules by gene transfer of a putative peptide transporter. Nature 351, 323–324.
Stet, R.J., van Erp, S.H., Hermsen, T., Sultmann, H.A., Egberts, E., 1993. Polymorphism and estimation of the number of MhcCyca class I and class II genes in laboratory strains of the common carp (Cyprinus carpio L.). Dev. Comp. Immunol. 17 (2), 141–156.
Theodoratos, A., Whittle, B., Enders, A., Tscharke, D.C., Roots, C.M., Goodnow, C.C., Fahrer, A.M., 2010. Mouse strains with point mutations in TAP1 and TAP2. Immunol. Cell Biol. 88, 72–78.
Uebel, S., Kraas, W., Kienle, S., Wiesmuller, K.H., Jung, G., Tampe, R., 1997. Recognition principle of the TAP transporter disclosed by combinatorial peptide libraries. Proc. Natl. Acad. Sci. U.S.A. 94, 8976–8981.
Vos, J.C., Spee, P., Momburg, F., Neefjes, J., 1999. Membrane topology and dimerization of the two subunits of the transporter associated with antigen processing reveal a three-domain structure. J. Immunol. 163, 6679–6685.
Walker, B.A., van Hateren, A., Milne, S., Beck, S., Kaufman, J., 2005. Chicken TAP genes differ from their human orthologues in locus organisation, size, sequence features and polymorphism. Immunogenetics 57, 232–247.
Walker, J.E., Saraste, M., Runswick, M.J., Gay, N.J., 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J. 1, 945–951.
Yan, G., Shi, L., Faustman, D., 1999. Novel splicing of the human MHC-encoded peptide transporter confers unique properties. J. Immunol. 162, 852–859.
Zdobnov, E.M., Apweiler, R., 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.

Supplementary data

Table S1

Oligonucleotide sequences.

Designation	Nucleotide sequence 5'- 3'	Use / Function
APv	GGC CAC GCG TCG ACT AGT ACT TTT TTT TTT TTT TTT TTV	cDNA synthesis;5'RACE;full cDNA id
AUAP	GGC CAC GCG TCG ACT AGT AC	5'RACE; full cDNA id
APv2	GAC TCA GGA CTT CAG GAC TTA GTT TTT TTT TTT TTT TTT V	cDNA synthesis;5'RACE;full cDNA id
AUAP2	GAC TCA GGA CTT CAG GAC TTA G	5'RACE; full cDNA id
T7	TAA TAC GAC TCA CTA TAG GGC GA	sequencing
SP6	CTA TTT AGG TGA CAC TAT AGA ATA C	sequencing
TAP1FW1	ACC GGG CGI RTG RCM GAC TGG	partial cDNA id
TAP1FW2	TGA GYY WKY TGA TGT GGT A	partial cDNA id
TAP1RV1	GTA CAR RAY ACI GAC CTT CA	partial cDNA id
TAP1RV2	GWC TCR CCS TCY TCI TTG GC	partial cDNA id
DLTAP1FW1	GCA GCT CGG TCG ACG ACT AA	full cDNA & gene id
DLTAP1FW2	CCG GTG GGG CAT CGT TAA AT	full cDNA & gene id/southern
DLTAP1FW3	GAG GCC ATC ACA GTT ATG	sequencing (full cDNA)
DLTAP1RV1	TGG TTG GCC TTA GCC AGT GA	1st 5'RACE (cDNA synthesis)
DLTAP1RV2	TGA ACC TTT GCA GCA ATA GT	1st 5'RACE
DLTAP1RV3	TGG TGG AAG TGT CCA GTG AG	1st 5'RACE
DLTAP1RV7	CGT ATG CGG CTC ATG GTG AC	2nd 5'RACE (cDNA synthesis)
DLTAP1RV8	GAC GTT GTA CAT GAG GTC AC	2nd 5'RACE
DLTAP1RV9	GCA GCA GTC AGT AGC GTC AT	2nd 5'RACE
DLTAP1RV10	ATG ACA CAT ATC GGC TCC AG	3rd 5'RACE (cDNA synthesis)
DLTAP1RV11	CAC CAC TGT CAC AAC ATA ACC	3rd 5'RACE
DLTAP1RV12	GAG TGC CAC CCC CAC AGT T	3rd 5'RACE/southern
DLTAP1RV13	ACT CCC CAG TGC CCT GCA AA	sequencing (full cDNA & gene)
DLTAP1RV14	CTG TGA GGA GAC ACG GAT TA	full gene id
DLTAP1RV15	GGC AAA TGA CAC CAG GTC TCC AC	sequencing (full gene)
DLTAP1RV16	ATT TTC AGG TGC CAA ACT GC	sequencing (full gene)
DLTAP1RV17	GGT ACT GGT CTT TGT AGC TTA G	sequencing (full gene)
DLTAP1RV18	TTG ACA GGT CCT TGA TGA AC	sequencing (full gene)
DLTAP1RV19	TTC GTT CTC TGT GTC CAA GTC	sequencing (full gene)
TAP2FW1	CTC TGT GAR AYR TTC ATC CC	partial cDNA id
TAP2FW2	AGT GCW GGY TGY MGI GGA GG	partial cDNA id
TAP2RV1	GAG ARS AGM ACR GGI TCC TG	partial cDNA id
TAP2RV2	ACT SAC ACA GGT GCT YTT ICC	partial cDNA id
DLTAP2FW1	GTA CGA GGA CGA CTG TCG GAG AA	full cDNA & gene id
DLTAP2FW2	ATG GCA GAC AAT ACA ACA ATC ATC C	full cDNA & gene id/southern
DLTAP2FW3	TAC TCT GTG CCA TCA GTG CC	sequencing (full cDNA & gene)
DLTAP2FW4	GTC AGC ACA GAG GGA AAA CT	sequencing (full cDNA & gene)
DLTAP2FW5	CTG TTT ATG AGA GTC TTA TAC	sequencing (full gene)
DLTAP2FW6	GTT CCT CGT ATT GAT AGA GAC	sequencing (full gene)
DLTAP2FW7	GGA GGC CCG TCG CTA TGA	sequencing
DLTAP2FW8	GAC AAC TGG CAG CCT GGT TTC	sequencing (full gene)
DLTAP2RV1	CTC TAG ACA ACC TGG ATG TG	5'RACE (cDNA synthesis)
DLTAP2RV2	TCA CCT GTC TTT ATG GTC TC	5'RACE
DLTAP2RV3	GGC ACT GAT GGC ACA GAG TA	5'RACE
DLTAP2RV5	ACT GTT GCA TAA AAC TTT TAT TCA TC	full gene id
DLTAP2RV6	ATT CAT CTT TAC CAG ATA CCA AGA C	full gene id
DLTAP2RV7	GCT CTC TGG CTT TCT GCT TCT	sequencing (full gene)
DLTAP2RV8	CAG AGG GGA AAA TCC ATC ATA C	sequencing (full gene)
DLTAP2RV9	TCT GCC GCC ACT AAC CAC A	sequencing (full gene)

Table S2

Accession numbers of sequences used in the present study.

Species	TAP1 (ABCB2)	TAP2 (ABCB3)	TAPL (ABCB9)
Homo sapiens (human)	GenBank: CAA40741	GenBank: NP_000535	GenBank: Q9NP78
Mus musculus (mouse)	GenBank: AAB41962	GenBank: AAA39609	GenBank: Q9JJ59
Rattus norvegicus (rat)	GenBank: CAA40742	GenBank: CAA53055; CAA45339	-
Mesocricetus auratus (Syrian hamster)	GenBank: AAB58721	GenBank: AAB58722	-
Gallus gallus (chicken)	GenBank: CAH58737	GenBank: CAH58738	-
Xenopus laevis (African clawed frog)	GenBank: AAP36718	GenBank: NP_001081860	-
Ginglymostoma cirratum (nurse shark)	GenBank: AAL59858	-	-
Salmo salar (Atlantic salmon)	GenBank: CAB05917	GenBank: ABQ01992; ABQ59650	-
Oncorhynchus mykiss (rainbow trout	GenBank: ABB52828	GenBank: AAB62237; AAD53035	-
Heterodontus francisci (horned shark)	GenBank: AF108387+ AF1088386	GenBank: CAH58738	-
Tetraodon nigroviridis (pufferfish)	GenBank: CAF88941	-	-
Danio rerio (zebrafish)	-	GenBank: CAK04959	GenBank: CAM46975
Oryzias latipes (Japanese medaka)	-	GenBank: BAD93260	-
Oryzias dancena (medaka)	-	GenBank: ACN49154	-

Table S3

Percentages of identity and similarity calculated with MatGAT.

		Similarity (%)	Identity (%)
TAP1	Dila-TAP1*01	-	-
	Dila-TAP1*02	99.9	99.7
	Dila-TAP1*03	99.7	99.6
	T. nigroviridis	73.8	59.1
	S. salar	84.3	67.8
	O. mykiss	84.4	69.1
	G. cirratum	64.8	46.3
	H. francisci	69.9	50.4
	X. laevis	69.4	47.4
	G. gallus	52.6	35.9
	R. norvegicus	63.6	40.9
	M. musculus	62.6	41.0
	M. auratus	62.6	40.1
	H. sapiens	62.7	40.9
TAP2	Dila-TAP2*0101	-	-
	Dila-TAP2*0202	95.5	95.5
	Dila-TAP2*02	99.6	99.2
	O. latipes	88.2	75.5
	O. dancena	88.3	75.5
	O. mykiss 2a	66.9	47.0
	O. mykiss 2b	82.0	66.3
	S. salar 2a	83.2	67.0
	S. salar 2b	81.5	66.0
	D. rerio	78.6	62.9
	H. francisci	68.9	46.9
	X. laevis	65.8	43.2
	G. gallus	60.6	41.6
	R. norvegicus 2a	64.1	41.9
	R. norvegicus 2b	63.1	41.8
	M. musculus	62.4	42.3
	M. auratus	62.7	42.1
	H. sapiens	63.6	41.5

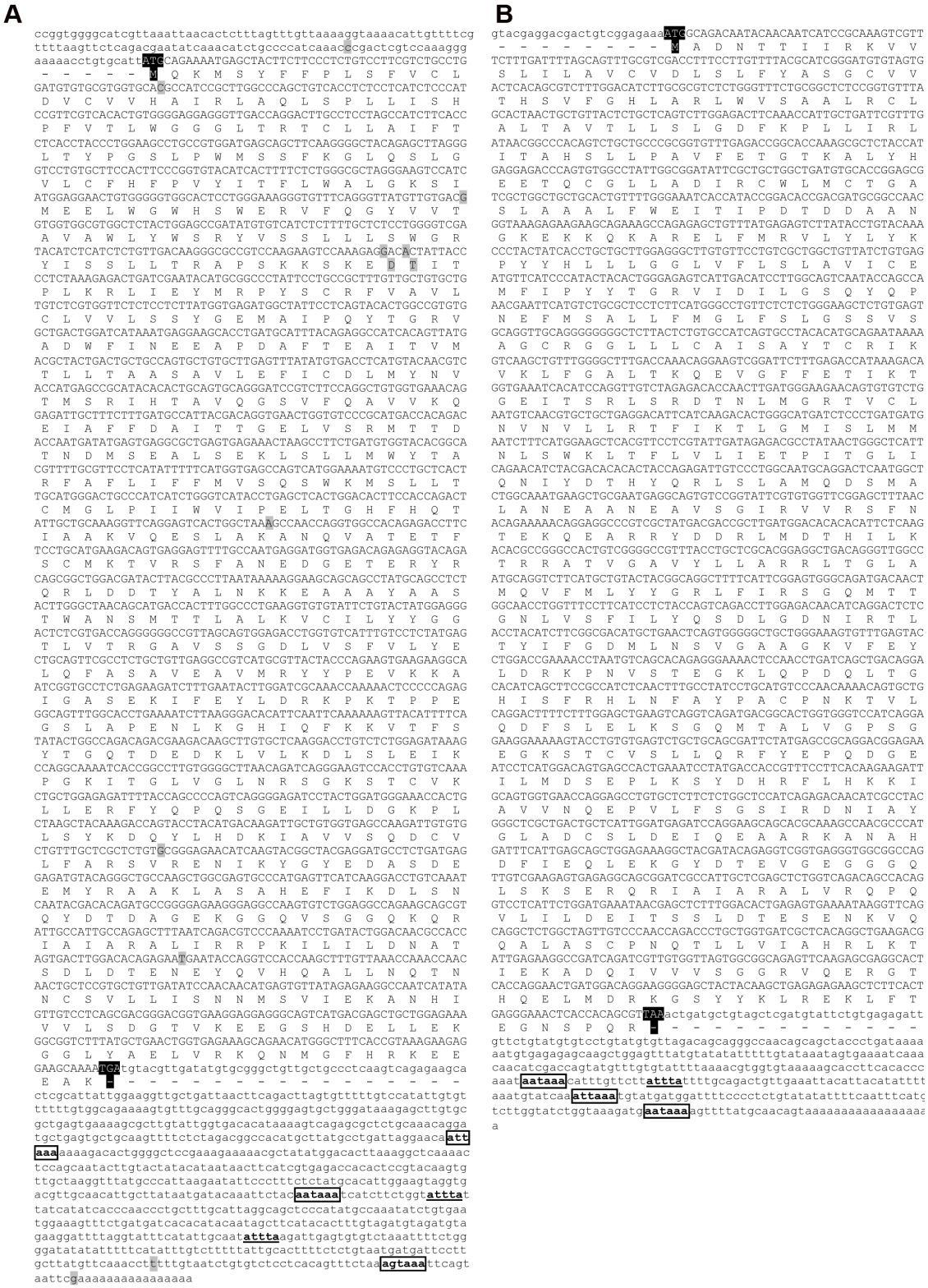
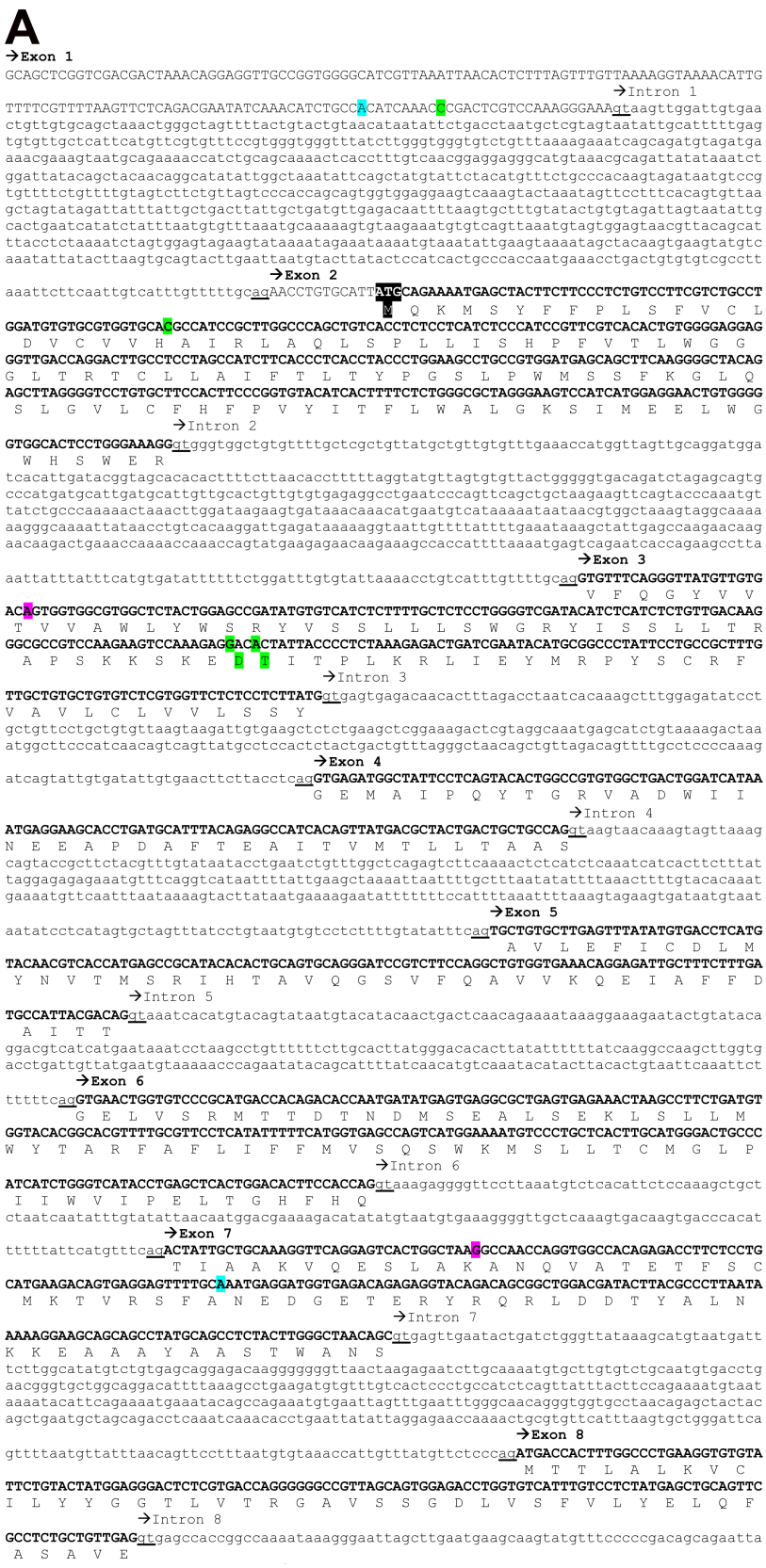


Figure S1.1. Nucleotide and deduced amino acid sequences of sea bass TAP1 (A) and TAP2 (B) cDNAs. The ORF and the predicted protein are shown in uppercase letters. The translation start and stop codons and initial methionine are shown in white over black background. Within the 3'UTR, the poly-adenylation signal sequences are highlighted in bold with boxes and mRNA instability motifs are bold and underlined. Differences between the Dila-TAP1 cDNA sequences (accession numbers [GenBank: HQ328079] and [GenBank: HQ328081]) are shaded in grey (c=t; c=t; g=a; g=c [H]; a=t [S]; a=g; g=c; t=c; t=g; g=a; and one extra c).



t=a; t=c; t=g [S]; a=t [A]; a=g; t=c; g=c; c=t).

Figure S1.2. Nucleotide and deduced amino acid sequences of sea bass TAP1 (A) and TAP2 (B) genes. The exons are shown in uppercase letters, while introns are shown in lowercase. The ORF is in bold. The translation start and stop codons and initial methionine are shown in inverted type. Intron donor/acceptor sites are underlined. Within the 3'UTR, the poly-adenylation signal sequences are in bold with boxes and mRNA instability motifs are bold and underlined. Differences to the cDNA sequences (accession numbers are [GenBank: HQ328080] and [GenBank: HQ328083]) are indicated: (a) cyan – residues different in both Dila-TAP1*01 and *02 (a=c; a=c; g=a; c=t [L]; t=c); green – residues different in Dila-TAP1*01 (c=t; c=t; g=c [H]; a=t [S]); pink – residues different in Dila-TAP1*02 (a=g; g=a; c=g; c=t; g=t). Residues shaded in grey within the 3'UTR denote the place of a 10-nucleotide insertion detected in both cDNA sequences (gaagtaggtg). (b) cyan – residues different in Dila-TAP2*01 (a=c; t=c; g=a; g=c [C]; a=g [V]; g=a [I]; c=g [Q];

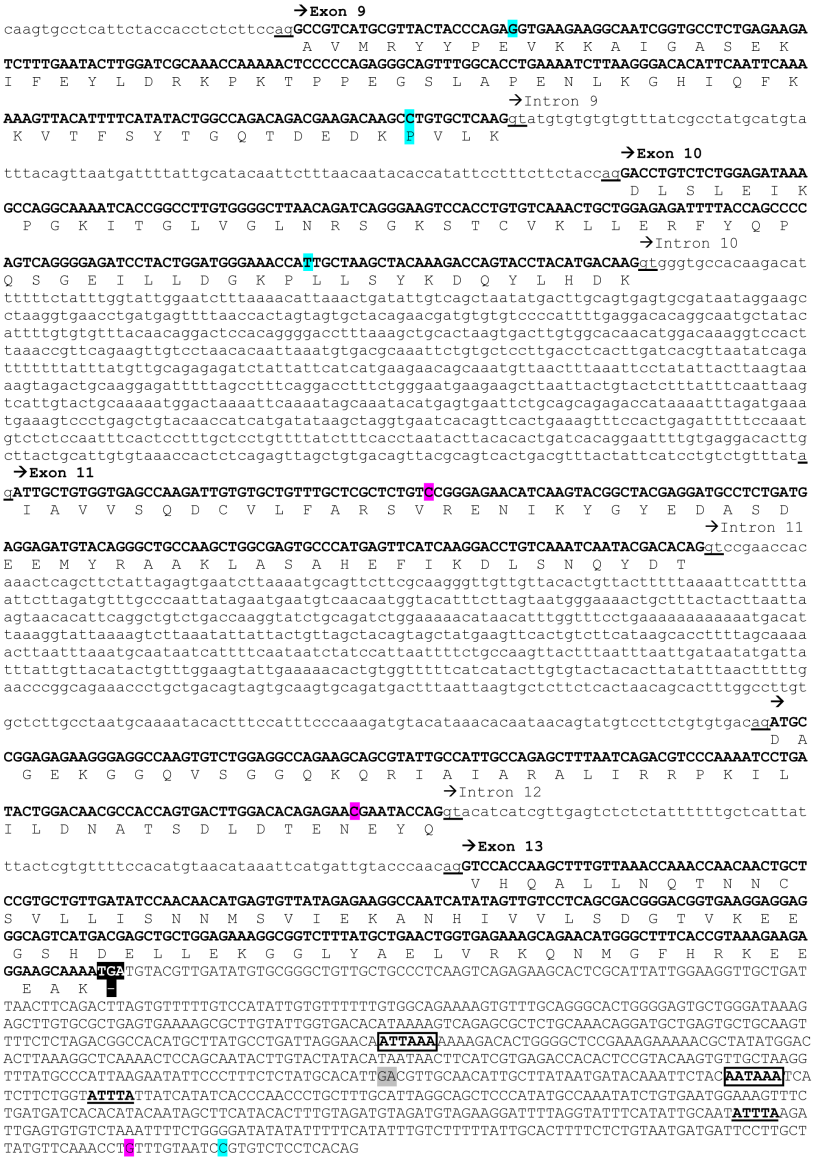


Figure S1.2 *(continued)*

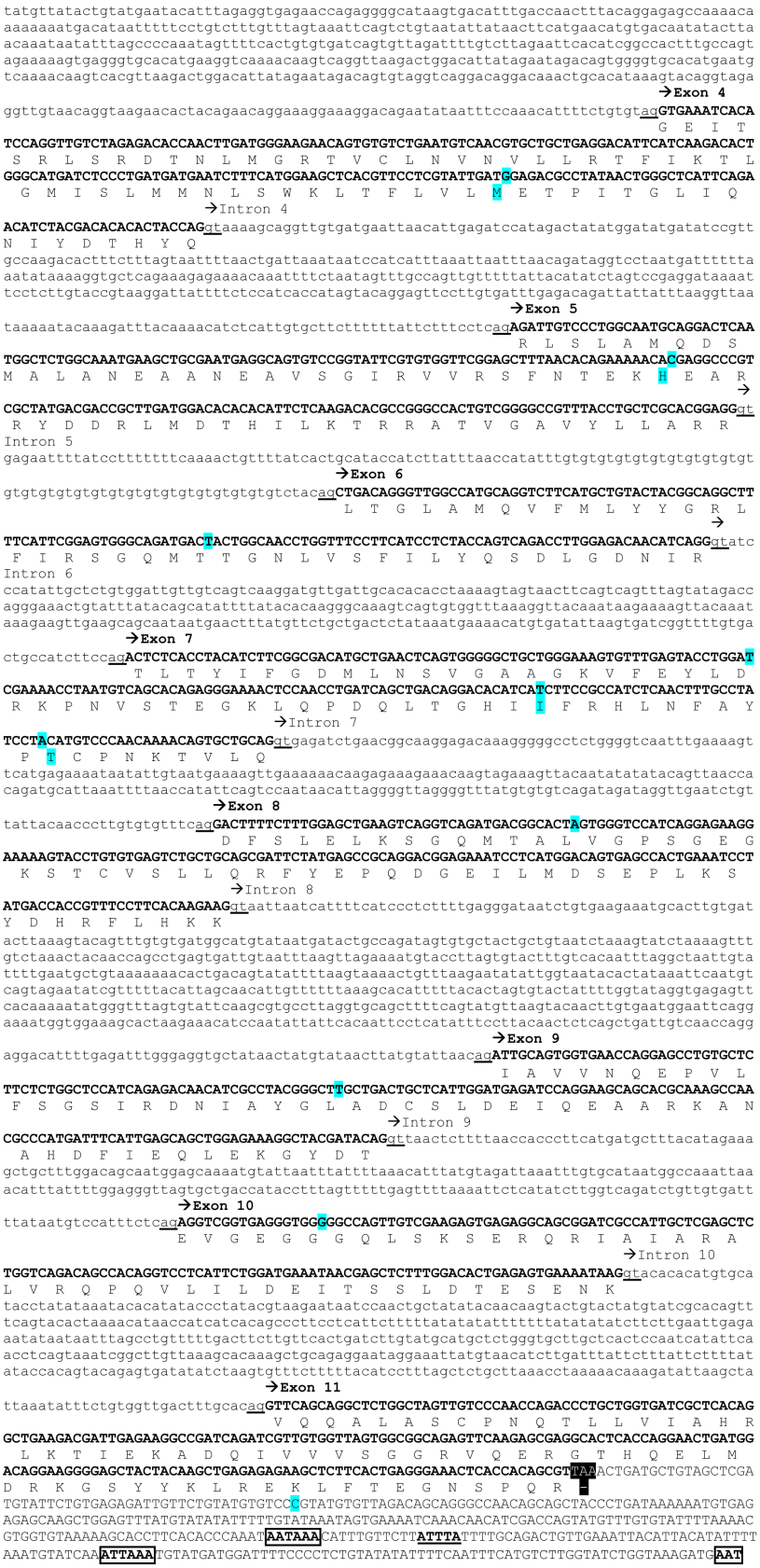
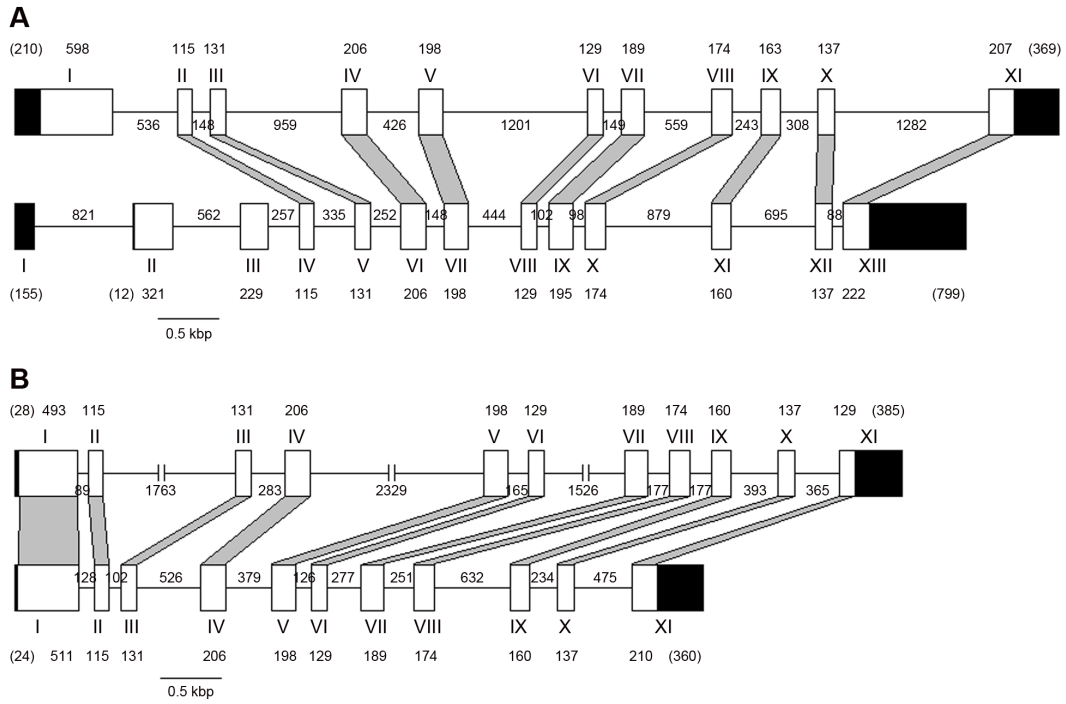


***Figure S1.2** (continued)*



Fgure S2. Schematic representation of sea bass and human TAP1 (A) and TAP2 (B) gene structures. The human gene is represented above the sea bass one. Boxes represent exons and horizontal lines introns. White and black boxes differentiate coding and untranslated regions, respectively. Exon and intron sizes are indicated in base pairs. Exon numbers are indicated by Roman numerals.

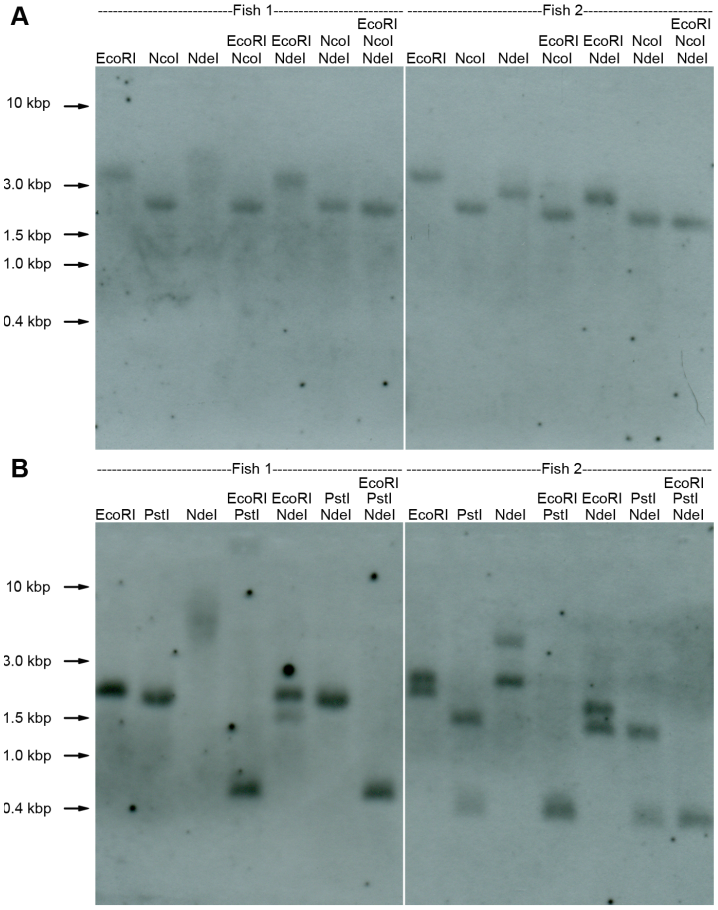


Figure S3. Sea bass TAP1 (A) and TAP2 (B) are both single copy genes. Southern blot analysis of genomic DNA from two animals digested with three restriction enzymes (isolated and combined). The blots were probed with a region of TAP1 or TAP2 (partial exon I). Sizes of the MW marker are indicated on the left.

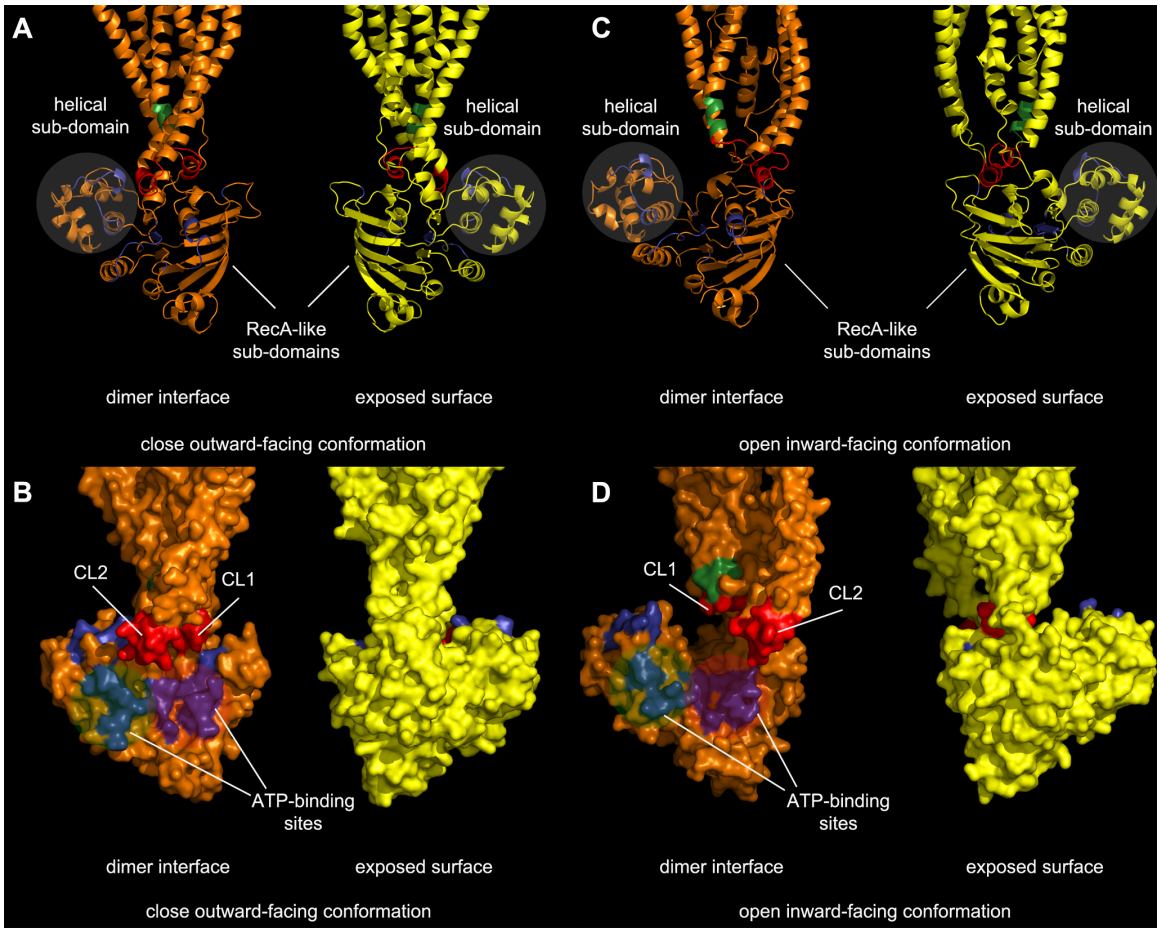


Figure S4. **3D homology models of accessory domain-less Dila-TAP1/-TAP2 dimer.** (A) Close-up representation of the modelled heterodimer in closed outward-facing conformation. The two subunits are shown side-by-side in the dimer-forming orientations, shifted along the x axis. Dila-TAP1 (orange) in NBD dimer interface view and Dila-TAP2 (yellow) in exposed surface view. Both RecA-like and helical sub-domains of the NBDs are labelled. The coupling loops are coloured red, the peptide sensor green and conserved motifs of the NBD blue. (B) Molecular surface representation of the Dila-TAP dimer in the same orientation as in panel A. Nucleotide-binding sites and coupling loops are indicated for Dila-TAP1. The colour code is the same as in A. (C) Close-up representation of the modelled heterodimer in open inward-facing conformation. The two subunits are shown side-by-side in the dimer-forming orientations, shifted along the x axis. The colour code is the same as in A. (D) Molecular surface representation of the Dila-TAP dimer in the same orientation as in panel C. Nucleotide-binding sites and coupling loops are indicated for Dila-TAP1. The colour code is the same as in A.

Mapping several residues on the 3D models

Several residues discussed in Section 3.3 were mapped on the 3D models of TAP1 and TAP2 and analysed in that context.

Of the predicted glycosylation sites, only N^{672}QT from TAP1 and N^{488}KT, and N^{660}QT from TAP2 are surface-exposed and hence can possibly be modified. Although exposed, N^{293}LS, is located very close to the ER membrane in both 2hyd- and 3g61-based models, and N^{458}VS is also not accessible in the 3g61-based one. The remaining TAP1 residues are probably not post-translationally modified, since they are facing the lumen of the translocation cavity (N^{234}), or located at the NBDs dimer interface (N^{653}, N^{685}). N^4TT was excluded from this analysis since it is located in the N-terminal accessory domain, for which no model could be calculated.

The mTAP2 T^{293}P corresponds to Dila-TAP2 T^{299} and is located in a helicoidal transmembrane region from the TMD (in both 2hyd- and 3g61-based structures), where replacement for a proline would probably disrupt the secondary structure and therefore alter the protein folding, which can possibly explain the observed reduction of protein levels for this mutant in mouse.

The hTAP2 A^{373}D corresponds to Dila-TAP2 A^{380} and is located in the contact interface between TAP1 and TAP2 helices, being oriented towards the lumen of the translocation cavity in the 3g61-based model, but not in the 2hyd-based one. The hTAP2 R^{380} corresponds to Dila-TAP2 R^{386}. This residue is also located at the wall of the translocation cavity with its side chain in the lumen in the 2hyd-based model, but not in the 3g61-based one. At least in one of the predicted TAP dimer structures, both mentioned residues can possibly have the same functions in sea bass (change the preference of bound peptide and influence transport activity, respectively).

The conserved residues hTAP1 K^{544}/ hTAP2 K^{509} correspond in sea bass to TAP1 K^{530} and TAP2 K^{515}. According to the 2hyd-based model, both these residues can putatively interact with the g-phosphate of the nucleotide as suggested in the previous section.

The aspartate from the D-loop has been reported to interact with the Walker A motif. Accordingly, the Dila-TAP1 D^{659} (D-loop) should interact with Dila-TAP2 N^{511} (Walker A), and Dila-TAP2 D^{644} (D-loop) with Dila-TAP1 S^{522} (Walker A). Both interactions can be maintained, since each aspartate side chain can bind to the main chain amide of the interacting residue. As in mammals (2ixf-based structure), an H-bond between the N^{511} OD1 and the D^{659} main-chain N can be formed.

The carbonyl of the D-loop alanine interacts with the side chain of the H-loop histidine. Given the nature of the interaction, it can be understood why this alanine is not conserved across species (aspartate, asparagine, cysteine, or serine in TAP1; serine or cysteine in TAP2). The H-loop is also variable in TAP1 molecules (asparagine, histidine, glutamine, arginine or glycine), but is conserved across species in TAP2 molecules (histidine). In sea bass, Dila-TAP1 D^{659} could interact with Dila-TAP2 H^{668} and Dila-TAP2 S^{642} with Dila-TAP1 N^{684}. This interaction probably occurs across species. In the TAP2 D-loop/TAP1 H-loop contact, the main chain carbonyl from the D-loop residue (alanine, serine or cysteine) interacts with side chain residue of the H-loop. In this case, asparagine (sea bass), histidine and glutamine in the H-loop can establish the contact; accommodation of the bulky arginine side chain would probably imply a rearrangement of this segment, while a glycine residue would be unable to establish this contact.

In the C-loop (LSGGQ), mutation of serine to alanine decreases ATPase activity. This residue is conserved in TAP1 molecules across species, but some TAP2 molecules naturally bear an alanine at this position, namely pufferfish and mammals. In sea bass, Dila-TAP1 S^{629} probably hydrogen bonds to Dila-TAP2 S^{522} (Walker A) in a side chain to side chain interaction. Furthermore, according to the 2ixf-based model, both serines are in position to interact with the ATP g-phosphate. In mammalian TAP2, the Walker A serine is replaced by an asparagine (N^{505}), therefore the interaction is also viable. Mutation of the C-loop serine to an alanine results in the loss of the interactions in the nucleotide-binding pocket, maintained across species. In the 2hyd-based model, Dila-TAP2 S^{614} may interact with Dila-TAP1 N^{511} (Walker A) also through a hydrogen bond. In mammals, the alanine from the C-loop should interact with an asparagine in the Walker A. The longer asparagine side chain can probably compensate for the lack of the side chain hydroxyl in its interaction partner, making a contact to the alanine main chain in TAP2. In all other species this interaction is most likely lost. However, in the 2ixf-based model, this interaction is no longer plausible.

Of the conserved residues described to interact with the nucleotide, only the Q-loop Q^{586} (Dila-TAP1 Q^{572} and Dila-TAP2 Q^{557}) does not interact with ADP in the 2hyd-based model neither with ATP in the 2ixf-based model. The other three contacts are plausible in the models: Y^{512} (A-loop), V^{520}, and D^{667} (Walker B) (Dila-TAP1 Y^{496}, V^{506}, D^{652}; and Dila-TAP2 Y^{481}, V^{491}, D^{637}). As in mammals, D^{652} presumably binds one of the water molecules that coordinate the Mg^{2+}.

CHAPTER 4

Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) tapasin

Rute D. Pinto, Diogo V. da Silva, Pedro J. B. Pereira and Nuno M. S. dos Santos

Fish & Shellfish Immunol. 2012 32:110-120. doi:10.1016/j.fsi.2011.10.029

Fish & Shellfish Immunology 32 (2012) 110–120

Contents lists available at SciVerse ScienceDirect

Fish & Shellfish Immunology

journal homepage: www.elsevier.com/locate/fsi

Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) Tapasin

Rute D. Pinto [a,c], Diogo V. da Silva [a], Pedro J.B. Pereira [b], Nuno M.S. dos Santos [a,*]

[a] *Fish Immunology and Vaccinology Group, IBMC – Instituto de Biologia Molecular e Celular. Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal*
[b] *Biomolecular Structure Group, IBMC – Instituto de Biologia Molecular e Celular. Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal*
[c] *ICBAS – Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Largo Prof. Abel Salazar 2, 4099-003 Porto, Portugal*

ARTICLE INFO

Article history:
Received 10 February 2011
Received in revised form
21 October 2011
Accepted 24 October 2011
Available online 18 November 2011

Keywords:
Dicentrarchus labrax
Tapasin
Gene cloning
Comparative sequence analysis
Protein structure prediction

ABSTRACT

Mammalian tapasin (TPN) is a key member of the major histocompatibility complex (MHC) class I antigen presentation pathway, being part of the multi-protein complex called the peptide loading complex (PLC). Several studies describe its important roles in stabilizing empty MHC class I complexes, facilitating peptide loading and editing the repertoire of bound peptides, with impact on $CD8^+$ T cell immune responses. In this work, the gene and cDNA of the sea bass (*Dicentrarchus labrax*) glycoprotein TPN have been isolated and characterized. The coding sequence has a 1329 bp ORF encoding a 442-residue precursor protein with a predicted 24-amino acid leader peptide, generating a 418-amino acid mature form that retains a conserved N-glycosylation site, three conserved mammalian tapasin motifs, two Ig superfamily domains, a transmembrane domain and an ER-retention di-lysine motif at the C-terminus, suggestive of a function similar to mammalian tapasins. Similar to the human counterpart, the sea bass TPN gene comprises 8 exons, some of which correspond to separate functional domains of the protein. A three-dimensional homology model of sea bass tapasin was calculated and is consistent with the structural features described for the human molecule. Together, these results support the concept that the basic structure of TPN has been maintained through evolution. Moreover, the present data provides information that will allow further studies on cell-mediated immunity and class I antigen presentation pathway in particular, in this important fish species.

1. Introduction

MHC class I molecules exist on the surface of almost all nucleated cells, and typically present intracellular, self and non-self peptides to be screened by the T-cell receptor on $CD8^+$ T lymphocytes [1]. Only foreign peptides should trigger cytotoxic T cell activation, whose action would culminate in the death of the cell presenting the antigen. The class I complexes are formed by the heavy chain (HC), light chain β2-microglobulin (β2m) and peptide [2]. Antigen processing is a tightly regulated process that occurs in the endoplasmic reticulum (ER), with assistance of several resident chaperones [3]. A lot of effort has been devoted to studying the class I presentation pathway, allowing to identify several important steps: generation of peptides in the cytosol; transport of peptides into the ER; initial folding of the HC upon entry into the ER, association with β2m and incorporation in the peptide loading complex (PLC); loading of peptide; exit from the ER and migration to the cell surface [4]. Cytosolic ubiquitinated proteins are degraded in the cytoplasm by the proteasome [5]. Some of the originated peptides are transported into the ER through the transporter associated with antigen presentation (TAP1/TAP2) [6] and once inside they are further trimmed by aminopeptidases (ERAP1/ERAP2) [7,8]. Class I HCs are glycoproteins, which initial folding is assisted by ER resident calnexin and also by ERp57 [9,10]. Only then the HC associates with β2m to form the heterodimer that being highly unstable needs further assistance by other chaperones: calreticulin, ERp57, tapasin and TAP heterodimer form the PLC which stabilizes the empty dimer and helps to load high affinity peptides [11]. Tapasin is a central molecule in this process [12] since it makes the bridge between the class I itself and the transporter [13]; without tapasin (TPN, also called TAPBP for TAP binding protein) class I molecules are unable to bind TAP and surface expression is impaired [14]. So on one hand, TPN maintains the structure of class I prior to peptide loading [15], and on the other hand it stabilizes and activates TAP

* Corresponding author. Tel.: +351 226074900; fax: +351 226099157.
E-mail addresses: rsp@ibmc.up.pt (R.D. Pinto), diogosilva85@gmail.com (D.V. da Silva), ppereira@ibmc.up.pt (P.J.B. Pereira), nsantos@ibmc.up.pt (N.M.S. dos Santos).

1050-4648/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.fsi.2011.10.029

to increase peptide supply [16–18]. Recently, besides these cell biological effects, evidence for the biochemical effect of TPN on peptide editing, by accelerating peptide exchange, has been reported [15,19]. TPN also binds to ERp57 trough a covalent bond [20], this way protecting the empty peptide-binding groove against α2 disulfide reduction, until that protection is provided by a peptide ligand [21]. Considered the functional unit of the PLC for its role in stabilizing empty MHC class I complexes, in facilitating peptide loading and in editing the repertoire of bound peptides [22], the crystal structure of the TPN/ERp57 conjugate has been recently published, outlining important residues in the contact of these two molecules and also in the contact to the HC through mutagenesis studies [23]. Other interactions between the PLC components include the interaction of calreticulin with ERp57 [24], and the lectin-like association of calreticulin with the monoglycosylated N-linked glycan on the HC [13,25]. According to two independent groups, PDI (protein disulfide isomerase) also associates with the PLC [26,27]. Recently, Kim et al. [4] proposed a hypothetical model for the role of PDI, where after catalysing oxidation of the α2 disulfide bond, PDI transfers the peptide it is carrying to the MHC class I molecule.

Like other members of the antigen presentation pathway, the TPN gene relies within the MHC: in mammals [28,29] and in birds [30], in the extended class II region, and in bony fish [31–34] in the class I region, being part of a true class I cluster in these ancient vertebrates. Furthermore, as other class I genes, TPN has been shown to be IFN-inducible in mammals [35] and also in trout [36] and salmon [37].

In sea bass, little is known about the MHC class I genes. The need to understand and characterize the antigen processing and presentation pathway that could bring insights for further studies of specific cell-mediated immunity in sea bass, served as the starting motivation for the present study. At that time, it was encouraged by the fact that homologues of mammalian TPN had been discovered in few teleost species [34,36–38], however none from the perciforms. In this work, we identify and characterize the TPN from sea bass, an important species to the aquaculture industry. For the first time we describe its transcript and gene and perform a comparative analysis at the molecular, structural and phylogenetic levels using *in silico* approaches, e.g. homology modelling.

2. Materials and methods

2.1. Ethics statement

This study was carried out in accordance with European and Portuguese legislation for the use of animals for scientific purposes (Directive 86/609/EEC; Decreto-Lei 129/92; Portaria 1005/92). The Portuguese competent authority for animal protection, Direcção Geral de Veterinária, approved the work (approval ID: 520/000/000/2006).

2.2. Fish

Sea bass, *Dicentrarchus labrax*, were kept in a recirculating, ozone-treated salt-water (25–30‰) system at 22 ± 1 °C, and fed with commercial pellets, twice a day. Fish were sacrificed with a lethal dose of 2-phenoxyethanol (Panreac; >5 mL/10 L).

2.3. cDNA cloning

Total RNA previously obtained from the head kidneys of three stimulated fish, and extracted according to the Versagene™ RNA tissue kit (Gentra Systems) protocol, was reverse transcribed following the instructions of the BioScript RNase H⁻ (Bioline) protocol, with primer APv2 (Table 1). Degenerate primers were designed based on conserved regions from the multiple alignment of tapasin amino-acid sequences from *Danio rerio* (zebrafish, NP_571049), *Oncorhynchus mykiss* (rainbow trout, AAZ66039), *Salmo salar* (Atlantic salmon, AB227285), and *Oryzias latipes* (Japanese medaka, BAB83851). The cDNA was amplified with primers TPNFW1 and TPNRV1 (Table 1) in a PCR reaction adjusted to a final volume of 50 μl, containing 200 μM of each dNTP, 1.5 mM $MgCl_2$, 1× PCR buffer (Bioline), 0.4 μM of each primer, 2 μL of cDNA (1:20) and 1.25 U of *Taq* polymerase (Bioline). PCR cycling conditions were: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 54 °C for 30 s, 72 °C for 30 s; and 72 °C for 5 min. The obtained product of

Table 1
Oligonucleotide sequences.

Designation	Nucleotide sequence 5′→3′	Use/function
APv	GGC CAC GCG TCG ACT AGT ACT TTT TTT TTT TTT TTT TTV	cDNA synthesis; 5′RACE; full cDNA id
AUAP	GGC CAC GCG TCG ACT AGT AC	5′RACE; full cDNA id
APv2	GAC TCA GGA CTT CAG GAC TTA GTT TTT TTT TTT TTT TTT V	cDNA synthesis; full cDNA id
AUAP2	GAC TCA GGA CTT CAG GAC TTA G	Full cDNA id
T7	TAA TAC GAC TCA CTA TAG GGC GA	Sequencing
SP6	CTA TTT AGG TGA CAC TAT AGA ATA C	Sequencing
TPNFW1	TTC TGG AGT GCT GGT TTG T	Partial cDNA id
TPNRV1	CTC CAG STC CAY IGC CAC CTG	Partial cDNA id
DLTPNFW2	TGA AAA CCA ACG ATC CAA ACT CTG	Full cDNA & gene id; sequencing full gene
DLTPNFW3	CGA GCA GCA CAT CGC AGC AGA G	Full cDNA id
DLTPNFW4	ACT GGA CTG CGG CTT CTG G	Sequencing full cDNA & gene; southern
DLTPNFW5	TGG GTC GAG GAG GAG GTT TA	Sequencing full gene
DLTPNFW6	CCC CAT ATA CCT GCA AGC TG	Sequencing full gene
DLTPNFW7	TGG AGC TGA GCT GGG AGT TT	Sequencing full gene
DLTPNFW8	ATG GTT GGC GTA GCG CTG GT	Sequencing full gene
DLTPNRV1	AGA GAG GGG TGG CAG AAA GT	5′RACE (cDNA synthesis);sequencing full gene
DLTPNRV2	CAG CTG GGT CGG TAA TAA AG	5′RACE
DLTPNRV3	ACT CTG TCA GGG TTG ATG TC	5′RACE
DLTPNRV4	TCT GCA CTG TGG CTG TGT	Sequencing full gene
DLTPNRV7	GAG AAA TGT GTT TAT TTC ACT TA	Full gene id
DLTPNRV8	AAC AAT CCA GAC TTT GCT TAT GA	Sequencing full gene
DLTPNRV9	TTC CAG TGT TCA CGA GAA AAT AC	Sequencing full cDNA & gene
DLTPNRV10	CTT GCC ATC GTA AGC CAG AA	Sequencing full gene
DLTPNRV11	TTA GCG GTG TCA AGC TCC AG	Sequencing full gene; southern

Letter nucleotide code: V = A, C, G; Y = C, T; S = G, C; I = deoxyinosine.

~780 bp was purified (QIAquick Gel Extraction Kit, QIAGEN) prior to cloning into *Escherichia coli* XL-1 competent cells following the pGEM®-T Easy Vector Systems (Promega). Plasmid DNA was extracted and sequenced (MWG) with universal primers T7 and SP6 (Table 1).

The obtained partial nucleotide sequence from sea bass tapasin was used to design three specific reverse primers in order to obtain the 5′ untranslated region (UTR) of the molecule: DLTPNRV1, DLTPNRV2, DLTPNRV3 (Table 1). The previously isolated RNA was used to synthesize first-strand cDNA with primer DLTPNRV1 (Table 1), according to BioScript RNase H^- (Bioline) protocol. The cDNA was diluted to 150 μL in distilled water and loaded on a Sephadex (in H_2O) G-50 span column. The 5′ RACE System for Rapid Amplification of cDNA Ends, Version 2.0 (Invitrogen) was followed and 80 units of Recombinant Terminal Transferase (Roche) were used in order to dATP tail 21 μL of the purified cDNA. The tailed cDNA (5 μL) was first amplified with the primers APv and DLTPNRV2 (Table 1). The reaction was performed in the already described conditions but with annealing at 55 °C. A second PCR was performed with primers AUAP and DLTPNRV3 (Table 1) in the previously described conditions but with annealing at 57 °C. A PCR product of ~350 bp was obtained and later purified, cloned and sequenced as described above.

To obtain the full sea bass tapasin cDNA, specific primers were designed at the beginning of the 5′ UTR region. DLTPNFW2 (Table 1) and AUAP2 were used in a first amplification PCR with the reaction conditions previously described but adding 0.15 U of *Pfu* DNA polymerase (Promega) and using a different cDNA as template (spleen cDNA, previously available in the laboratory, from a stimulated fish, synthesized with APv2). The cycling conditions were as described before, but with annealing at 59 °C for 1 min, and extension for 1 min 30 s. Second amplification PCRs were performed using nested primer DLTPNFW3 (Table 1) and AUAP2. PCR conditions were the same as the just described above, except that annealing was at 60 °C for 30 s, and 1 μL of first amplification PCR product was used as template. The obtained ~2200 bp PCR products were purified, cloned and four independent clones were sequenced as before with primers T7, SP6, DLTPNFW4 and DLTPNRV9 (Table 1).

2.4. Genomic DNA cloning

Genomic DNA was isolated from sea bass erythrocytes from two different fish, as described by Stet et al. [39].

To obtain the full tapasin gene, genomic DNA from two different fish was amplified by PCR using primers DLTPNFW2/DLTPNRV7 (Table 1). The PCR was performed in a 50 μl PCR reaction (200 μM of each dNTP, 1.5 mM $MgCl_2$, 5 μL 10× PCR buffer (Bioline), 0.4 μM of each primer and 1.25 U of *Taq* DNA polymerase (Bioline) and 0.15 U of *Pfu* DNA polymerase (Promega)). The cycling conditions were: 94 °C for 2 min; 30 cycles of 94 °C for 45 s, 52 °C for 30 s, 72 °C for 3 min; and 72 °C for 5 min. The PCR products were purified, cloned and independent clones sequenced as before. For primers see Table 1.

2.5. Sequence analysis

Full nucleotide and protein sequences from sea bass tapasin were compared to several tapasin sequences currently available in the GenBank database retrieved using the BLAST program [http://www.ncbi.nlm.nih.gov/genbank/]. The multiple alignment was made using CLUSTALW [40] and formatted with Bioedit [41]. The putative leader sequence was predicted with SignalP 3.0 [42]. The Ig-like domains were based on InterProScan predictions [43]; and the transmembrane domain was based on hydrophobicity tendency and clusters predicted by ProtScale [http://www.expasy.org/tools/protscale.html] and DrawHCA [http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA], respectively, and also on secondary structure (α-helices) predicted by NetSurfP [44]. Molecular weight was calculated with Expasy compute pI/Mw tool [http://www.expasy.org/tools/pi_tool.html]. Possible N-glycosylation sites were predicted on Expasy post-translational modification tool [http://www.cbs.dtu.dk/services/NetNGlyc/]. The Neighbour-joining phylogenetic tree was constructed using MEGA version 3.1 [45], with p-distance parameter and complete deletion of gaps. The phylogenetic tree was tested for reliability using 1000 bootstrap replications. The percentages of similarity and identity were calculated with MatGAT [46] using default parameters.

2.6. Southern blotting

The gDNA was also digested with different restriction enzymes, separated in 0.8% agarose gel and subjected to Southern blotting [47]. Restriction enzymes EcoRI, NcoI, and NdeI (all zero cutters within the probe and the genomic region of the probe) isolated and combined were used to digest ~9 μg of gDNA overnight at 37 °C.

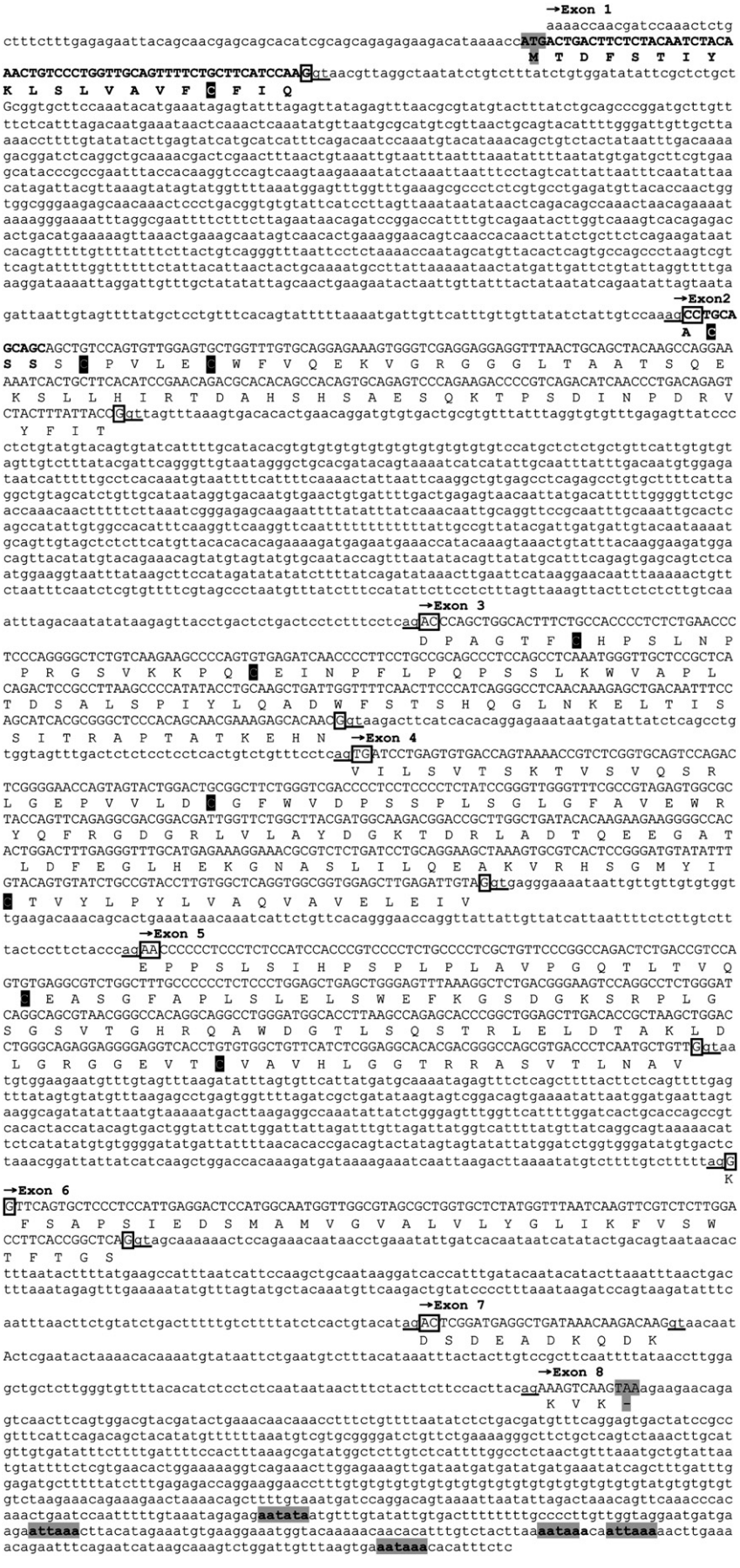
A portion of the sea bass tapasin cDNA, that included partial exons 4 and 5, was amplified with primers DLTPNFW4/DLTPNRV11 (Table 1) in the same PCR conditions (annealing at 59 °C) and later purified, labelled and used as a probe.

Preparation of the labelled probe, hybridization and post-hybridization stringency washes were performed accordingly to Gene Images™ AlhPhos Direct™ Labelling and Detection System (Amersham Biosciences) kit. For signal generation and detection the Chemiluminescent Signal Generation and Detection with CDP-*Star*™ protocol from the same kit was followed.

2.7. Homology modelling

A suitable structural template for Dila-TPN, human TPN [PDB: 3f8u] was identified by a BLAST search as implemented in the SWISS-MODEL Protein Modelling Server [48]. A manually edited sequence alignment (~30% identity between target and template, with gaps introduced essentially outside the secondary structure elements) was used for homology modelling with SWISS-MODEL [48] in alignment mode. The resulting theoretical model of the protein had a calculated global energy of −8666.288 kJ/mol, which was further minimized with PHENIX [49] to a total energy of −9308.911 kJ/mol. The differences between target and template rely essentially within the N-terminus of the protein where similarity is lower between the molecules and where there are two poorly ordered regions in the template. The secondary structure elements were compared using PROMOTIF [50]. Model quality was assessed with WHATCHECK [51] and PROCHECK [52] as implemented in SWISS-MODEL. Model figures were prepared with PyMOL [http://pymol.org/].

Fig. 1. Nucleotide and deduced amino acid sequences of sea bass TPN gene [GenBank: HQ328077]. The ORF and the predicted protein sequence are shown in uppercase letters. The translation start codon and respective residue and stop codon are shaded in grey. The predicted signal peptide is bold, and cysteines are shaded in black. Within the 3′UTR, putative poly-adenylation signal sequences are in bold with grey shading. The intron splicing consensus sites (gt/ag) are underlined. All introns except intron 7 interrupt the coding frame between the first and second bases, as in most members of the Ig superfamily (boxes).


→Exon 1
→Exon2
→Exon 3
→Exon 4
→Exon 5
→Exon 6
→Exon 7
→Exon 8

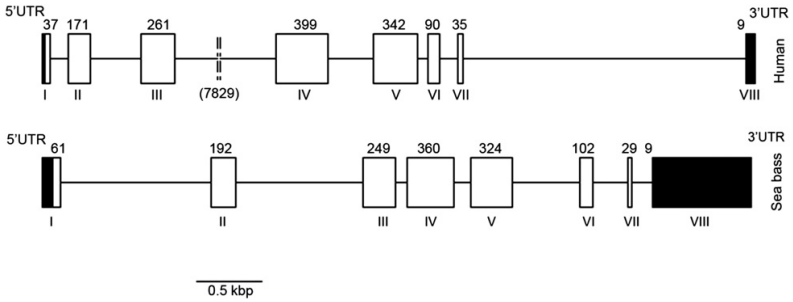


Fig. 2. Schematic representation of sea bass and human TPN gene structures. Boxes and horizontal lines represent exons and introns, respectively. White and black boxes differentiate coding and untranslated regions, respectively. Values above boxes represent the number of nucleotides. Exon numbers are indicated below boxes by Roman numerals.

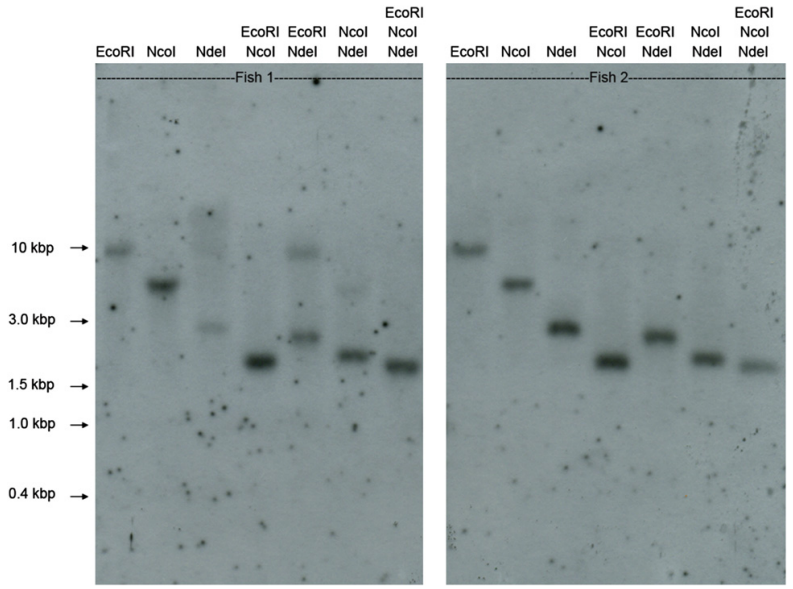


Fig. 3. Sea bass tapasin is a single copy gene. Southern blot analysis of genomic DNA from two animals digested with three restriction enzymes (isolated and combined). The blot was later probed with a region of tapasin that included parts of exons IV and V. All enzymes are zero cutters within the probe: one band can be seen for each of these individual digestions and for each set. Sizes of the Mw marker are indicated at left.

3. Results and discussion

Three full-length sea bass tapasin (Dila-TPN) cDNA sequences (Dila-TPN*0101 [GenBank: HQ328073], *0102 [GenBank: HQ328074], and *0103 [GenBank: HQ328075]) were obtained from the spleen of one fish and two full-length genomic sequences were obtained from erythrocytes of another two individuals (Dila-TPN*02 [GenBank: HQ328076], and Dila-TPN*03 [GenBank: HQ328077]).

The Dila-TPN gene (5496 bp) contains eight exons, and all intron donor/acceptor sites match the consensus motifs (Fig. 1A). Exon I encodes the 5′ UTR and most of the putative leader peptide, exon II and III the unique domain, exon IV the IgSf V domain, exon V the IgSf C domain, exon VI the putative hydrophobic transmembrane domain, and exons VII and VIII the cytoplasmic tail. Exon VIII also encodes the 3′ UTR. The codons in the exon boundaries are split by phase 1 introns (single exception of the phase 0 split at intron VII) (Fig. 1A), which together with the fact that some individual exons correspond to separate functional domains of the protein are typical features of members of the Ig gene superfamily [53]. However, disruption of the leader peptide by an intron, the presence of a lysine in the transmembrane region and the short cytoplasmic tail are atypical for class I or II loci and yet have also been described for human tapasin [28]. In general, the sea bass TPN genomic structure is very similar to those of human [28] (Fig. 2) and chicken [30] molecules except in the size of the introns: in Dila-TPN, introns 1 and 2 are slightly larger than those of human and chicken, while 3 and 7 are considerable smaller than the human counterparts.

The transcripts have a total size of 2193, 2195 or 2199 bp, with an open reading frame (ORF) of 1329 bp, a 5′UTR of 83 bp, and a 3′ UTR 781, 783 or 787 bp long (besides a polyadenylation tail) containing five potential polyadenylation signals. The different sizes of the 3′UTR are due to the presence of dinucleotide repeats with distinct sizes, $(GT)_{n=[2-30]}$, interrupted by one or two AT sequences: $(GT)_n[AT]_1(GT)_n$ or $(GT)_n[AT]_1(GT)_n[AT]_1(GT)_n$ (Fig. S1). The microsatellite is also present in the genomic sequences (Fig. S1). Although the pattern of repeats is difficult to interpret, it most likely results

from PCR/sequencing errors, especially considering that sea bass TPN is a single copy gene, as assayed by southern blotting (Fig. 3) and in accordance with what was described for zebrafish [38]. Other minor differences found when comparing the cDNA with genomic sequences or the two genomic sequences with each other are summarized in Tables S1–S3.

The 1329 bp ORF translates into a 442 amino-acid sequence (Fig. 4). Removal of the putative 24-amino-acid leader sequence of Dila-TPN, would result in a 418 amino-acid (~45 kDa) mature protein, making Dila-TPN slightly smaller than mammalian tapasins, which range in size from 428 (human) [14] and 443/444 (rat/mouse) amino acids [29,54], but similar to TPN of other fish, which contain 416–421 amino acids [34] (Fig. 4).

The ER luminal domains of TPN mediate interaction with the MHC class I HC and ERp57. The N-terminal region of sea bass TPN displays no homology to any conserved domain, as is the case for

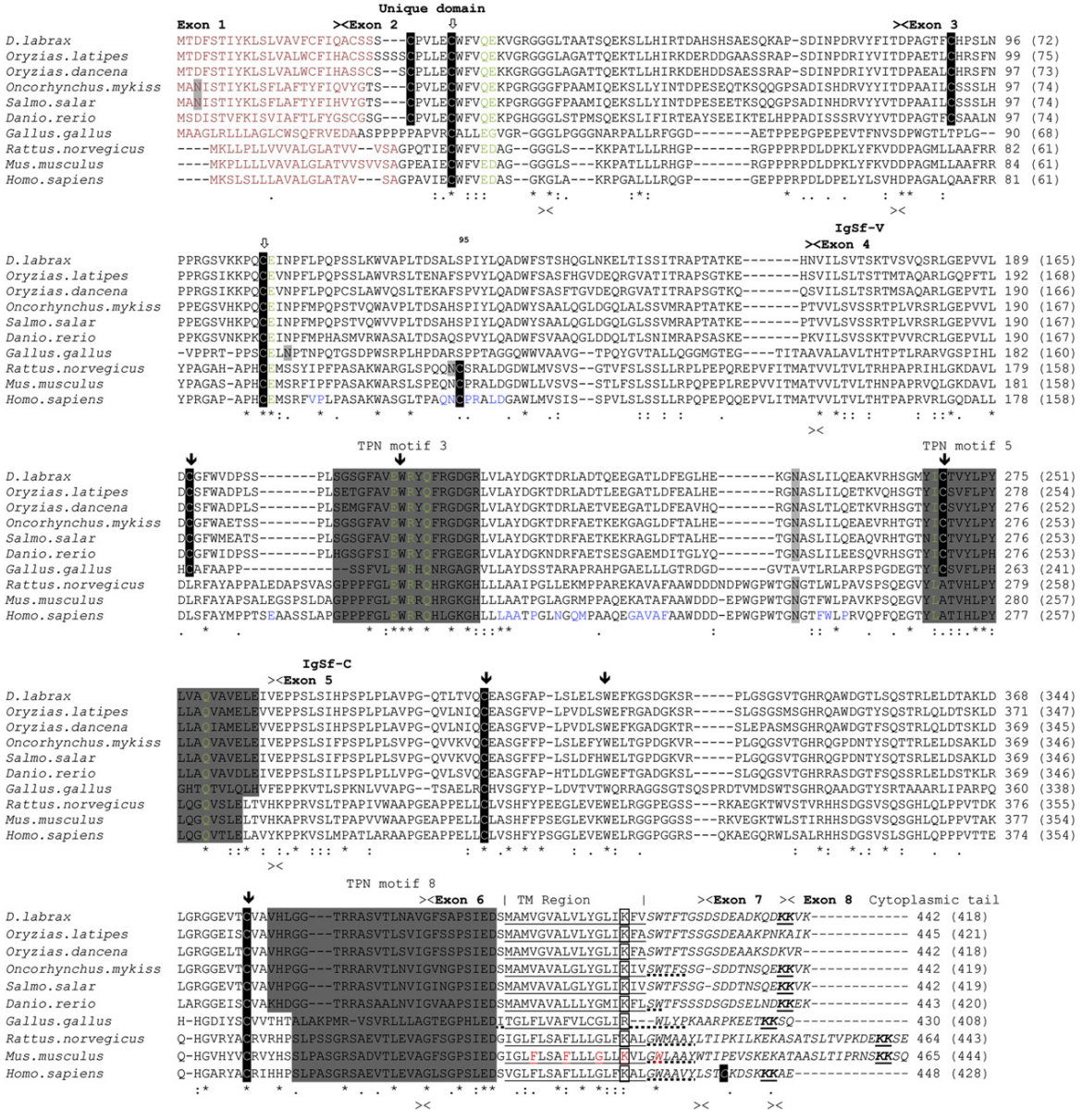


Fig. 4. Alignment of TPN amino acid sequences. The protein sequences of TPN from *Dicentrarchus labrax*, [GenBank: HQ328073]; *Oryzias latipes*, [GenBank: BAB83851]; *O. dancena*, [GenBank: ACN49160]; *Onchorynchus mykiss*, [GenBank: AAZ66039]; *Salmo salar*, [GenBank: AB227285]; *Danio rerio*, [GenBank: NP_571049]; *Gallus gallus*, [GenBank: O73895]; *Rattus norvegicus*, [GenBank: CAC34730]; *Mus musculus*, [GenBank: AAD17964]; *Homo sapiens*, [GenBank: AAC20076] were used in the multiple alignment. >< (above the alignment): sea bass exon/intron boundaries; >< (below the alignment): chicken/human exon/intron boundaries. Leader peptides are in dark red lettering. Tapasin unique domain, IgSf V-domain, and IgSf C-domain are indicatated above the alignment. The TM region is underlined, and the cytoplasmic tail is italicized. Three conserved tapasin motifs are shaded in dark grey, matching the tapasin motifs 3, 5 and 8 as established in mammals. Cysteines are shaded in black, and conserved cysteines within the unique domain of tapasin are indicated by open arrows. The cysteine involved in association with ERp57 is indicated by superscript 95. Conserved cysteine and tryptophan residues involved in the IgSf V and C domains fold are indicated by a solid arrow. Residues that affect binding to the MHC class I HC are highlighted in green. Residues within 4.5 Å of an ERp57 atom are labelled in blue. The lysine residue involved in association with TAP is boxed. Residues of mouse TMD tapasin that interact with TAP are coloured in red. ER-retention signals (di-lysine) are shown in bold type and underlined within the cytoplasmic region. N-glycosylation sites (N-X-S/T) are boxed in grey. Numbers at the right of the alignment denote sizes of the precursor and mature (within brackets) proteins. Dashes indicate gaps that maximize the alignment; (*) means identical residues, (:) and (.) denote conserved and semi-conserved substitutions, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

other organisms and so it is considered a unique tapasin region (Fig. 4). In particular, the 50 N-terminal amino acids of tapasin are required for efficient association with MHC class I HC [16]. Several conserved residues are found in this region, including two cysteines involved in an intra-chain disulfide bridge [23] (in sea bass C^{7}, C^{83}). By comparing tapasins from different species, Dong et al. [23] identified a prominent conserved patch on the surface of a broader N-terminal domain, which includes residues 1–269, encompassing the unique tapasin domain and the Ig V-like domain. Different mutations in this region led to reduced or abrogated binding of MHC class I HCs to the recombinant conjugate (TPN/ERp57). In this regard, of the eight mutated residues with the described effects on HC-binding, five are absolutely conserved (hE^{72}/sbE^{84}, hE^{185}/sbE^{185}, hR^{187}/sbR^{187}, hQ^{189}/sbQ^{189}, hQ^{261}/sbQ^{255}) and three are conservatively substituted (hE^{11}/sbQ^{11}, hD^{12}/sbE^{12}, hL^{250}/sbI^{244}), suggesting that these amino acids may also be important for TPN/MHC interaction in other vertebrates, including sea bass. It is also through the unique tapasin region that TPN interacts with ERp57 [55], namely through the disulfide bridge between TPN C^{95} and ERp57 C^{57}. However the tapasin cysteine (hC^{95}) involved in this covalent bond is only conserved in mammals, being replaced by a serine in all the other vertebrate species (sbS^{107}). Although previously postulated that this conjugate is preserved as a consequence of non-covalent interactions [20], and in spite of the recently characterized mouse tapasin C^{95} mutants found to participate in a non-covalent complex with ERp57, indicating that these molecules are able to interact in the absence of the disulfide bond in the mouse system [56], it remains to clarify how sea bass tapasin would associate with ERp57, as previously mentioned for other bony fish [36] and also for chicken [30], since only 8 of the 25 residues described to be in close proximity to ERp57 [23] are conserved (P^{90}, P^{108}, L^{111}, L^{198}, A^{199}, G^{214}, A^{215}, Y^{251}) and another 5 are conservatively substituted (L^{89}, S^{107}, Q^{112}, N^{205}, T^{216}) in sea bass (Fig. 4).

As expected, Dila-TPN retains the two Ig-like domains identified in tapasins from other organisms. Canonical cysteine and tryptophan residues (sb-C^{167}, sb-W^{186}, sb-C^{245} and sb-C^{288}, sb-W^{302} and sb-C^{353}), involved in disulfide bond formation and stabilization of the Ig fold [57], are present in the predicted IgSf V and C1 domains of sea bass TPN. The cysteines in the C domain are conserved across all species, but those in the V domain are absent in mammals, as others had previously reported [30,36,37]. Nonetheless, the Ig fold is still maintained in the C domain of human tapasin, as elucidated by the TPN-ERp57 crystal structure [23].

Within Dila-TPN exons IV, V and VI, 3 conserved motifs (matching motifs 3, 5 and 8) were identified (Fig. 4), of the total eight fingerprints as established from mammals that provide a signature for vertebrate tapasins. Within exon IV (Ig V domain)

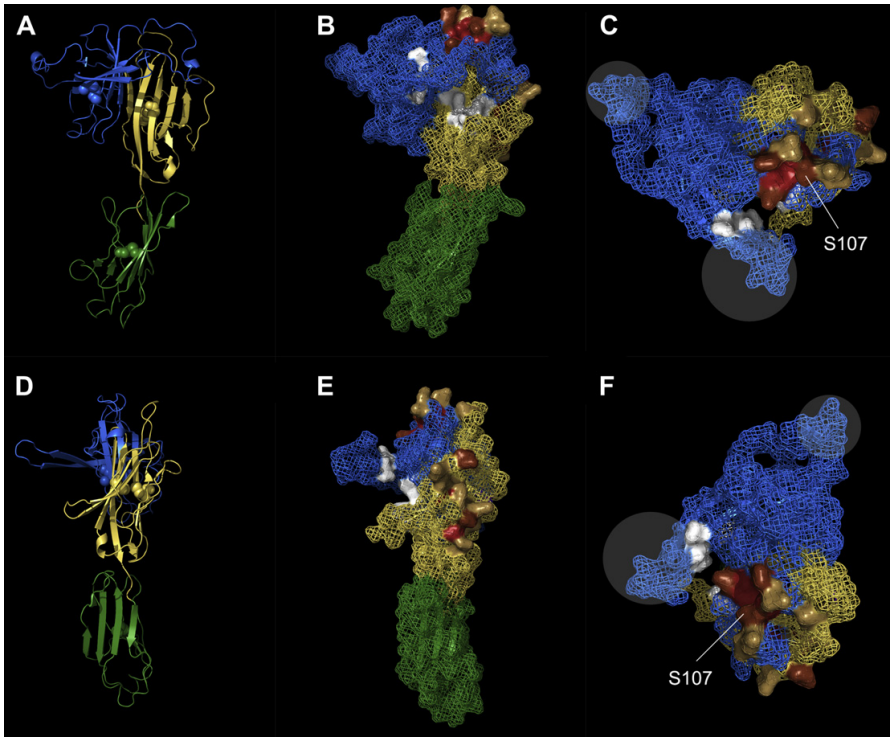


Fig. 5.1. Three-dimensional homology model of Dila-TPN (excluding the TM and cytoplasmic domains). Model based on the crystal structure of human TPN [PDB: 3f8u [23]]. (A) Representation of the Dila-TPN model with the N-terminal coloured blue, the Ig V-like domain yellow, and the Ig C-like domain green. This view of the molecule highlights its L-shaped structure. Spheres represent disulfide bridges. (B) Surface representation of Dila-TPN, in the same orientation as depicted in (A). Contacts to MHC class I heavy chain are labelled in white, and those to ERp57 are coloured red (conserved), brown (semi-conserved) and sand (non-conserved). (C) Surface representation of Dila-TPN rotated 90° clockwise around the *x*-axis in respect to the orientation presented in (B). Circles denote disordered regions on the template. (D) Representation of Dila-TPN rotated 90° counter clockwise around the *y*-axis in respect to the orientation presented in (A). (E) Surface representation of Dila-TPN, in the same orientation as depicted in (D). Colour code of highlighted residues as in (B). (F) Surface representation of Dila-TPN, rotated 90° clockwise around the *x*-axis in respect to the orientation presented in (E). Colour code of highlighted residues as in (B). Circles denote disordered regions on the template. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a single conserved putative N-glycosylation site (N^{227}) is preserved in sea bass tapasin. This particular residue is conserved across species, except for chicken that bears an alanine instead.

Although InterProscan failed to predict a transmembrane domain in Dila-TPN, hydrophobic-rich helical structures could be identified in this region using ProtScale, DrawHCA and NetSurfP. According to these predictions, the putative transmembrane region in the Dila-TPN molecule spans residues 380–402 of the mature protein, in agreement with the boundaries defined in the multiple sequence alignment in Fig. 4 (residues 382–398). TAP binding was suggested to occur within this region [14,16,18], namely through K^{408} (human), which is conserved in sea bass (K^{396}). This lysine, only absent in the chicken molecule, is important for proper MHC class I surface expression and in particular for tapasin stabilization of human TAP [58]. Although, *per se*, mutation of the equivalent residue of mouse tapasin (mTPN) did not abolish the ability to stabilize TAP2, mK^{431} and its helical neighbours (located in the same flank of the helix, mF^{420}, mF^{424}, mG^{428}, mW^{435}) are required for a fully functional TM domain [59]. Also mE^{414} in the connecting peptide is critical, but not sufficient for TAP2 stabilization [59]. All the above-mentioned residues (K^{396}, V^{385}, L^{389}, G^{393}, W^{400}, and E^{379} in sea bass) are conserved or conservatively substituted across species (Fig. 4), and importantly an mTPN chimera containing the zebrafish TM domain was able to stabilize mTAP2 [59]. Apparently, the mouse (but not the human) TPN TM/CYT domain is crucial to interactions involving the MHC class I HC as well as TAP, and is required for class I expression on the surface of mouse cells [60]. In the cytoplasmic tail, there is a double lysine motif, at positions −3 −4 (K^{415}/K^{416}), compatible with a retention signal in the endoplasmic reticulum [61], present in all sequences except those of medaka. According to Paulsson et al. [62], this is a retrieval signal important for the retrograde transport of TPN from the Golgi back to the ER, controlling the transport of unstable MHC class I molecules and hence the quality of MHC class I antigen presentation.

The Dila-TPN structure (except its transmembrane and cytoplasmic domains) was predicted by comparative modelling using the human tapasin crystal structure (PDB id: 3f8u [23]) as template. The generated homology model of Dila-TPN is compatible with the L-shaped structure of human tapasin, with one N-terminal domain consisting of a seven-stranded β-barrel, an intermediate and one C-terminal Ig-like domains (Fig. 5.1A, B, D, E). Furthermore, several important structural amino acids are conserved in the same spatial position (Fig. 5.1B, C, D, E). These include the pairs of bonded cysteines in the unique and IgSf C-like domains, the N-glycosylation site and all the residues postulated to interact with MHC class I HC. Also the residues in close proximity to ERp57 are in their majority structurally conserved. The additional pair of cysteines in the IgSf V-like domain is located so that a disulfide bond can be formed (Fig. 5.1A, D). Secondary structure elements (SSE) of the homology model and of the template were determined with PROMOTIF [50]. Regarding the unique domain, the model consists of two helices and seven β-strands. Compared to the template, helices H1 and H3 are missing due, in the case of the first helix, to the presence of a proline in sea bass, and there is one extra helix after strand S2, corresponding to insertions in the sea bass sequence (Fig. 5.2). Other differences include longer strands S1 and S2 (another region of insertions in the model) and shorter strands S3, S4, S6, S7 and S10, and helix H2. The main differences observed, within the Dila-TPN N-terminal domain between the target and the template, are due on one hand to two insertions (low similarity) in the model, and on the other hand correspondence to the two poorly ordered regions of the template (Fig. 5.1C, F). The loop connecting to the following domain is shorter in sea bass, representing deletions in the respective sequence (Fig. 5.2). Within the Ig V-like domain, the model displays

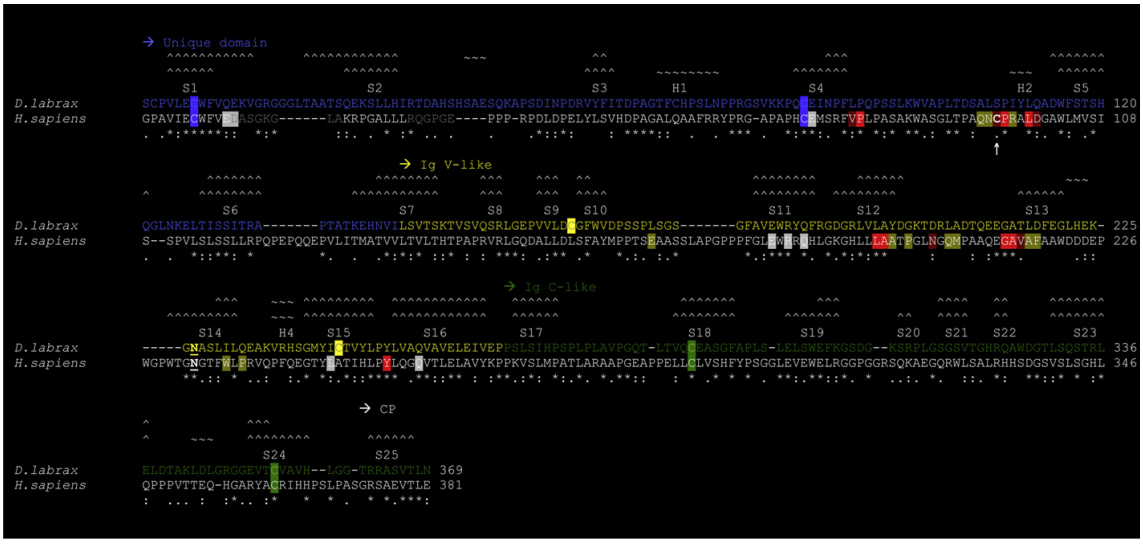


Fig. 5.2. Alignment of target and template amino acid sequences. The protein sequences of TPN from *Dicentrarchus labrax*, [GenBank: HQ328073] and *Homo sapiens*, [GenBank: AAC20076] were used in the multiple alignment. Secondary structure elements (SSE) are indicated above the alignment: (^) β-strand, (~) α-helix [determined with PROMOTIF [50] for sea bass (upper); based on crystal structure for human [23] (lower)]; numbering of SSE according to the human molecule. Tapasin unique domain, IgSf V-domain, and IgSf C-domain are denoted above the alignment and coloured blue, yellow and green, respectively. The two disordered regions of the template structure are in grey lettering. Cysteines involved in intra-chain disulfide bonds are shaded in blue, yellow or green according to the domain they belong to. Residues in close proximity to ERp57, including C^{95} (highlighted with an arrow below the alignment), are shaded in red (conserved), brown (semi-conserved) and sand (non-conserved). Mutated residues affecting MHCI binding to the conjugate are shaded in light grey. N-glycosylation sites (N-X-S/T) are in bold type and underlined. Dashes indicate gaps that maximize the alignment; (*) means identical residues, (:) and (.) denote conserved and semi-conserved substitutions, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

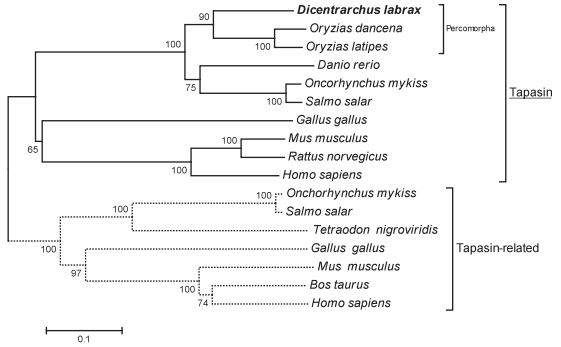


Fig. 6. Phylogenetic tree of TPN. An amino acid alignment (CLUSTALW) of the proteins was used to generate the unrooted Neighbour-joining tree. Node values represent percentages of bootstrap analysis from 1000 replications and complete deletion of gaps. Accession numbers from TPN sequences have been referred in Figure 4. Those of TPN-R, used as out-group, are as follows: *Homo sapiens* [GenBank: Q9BX59], *Bos Taurus* [GenBank: A6QNK8], *Mus musculus* [GenBank: Q8VD31], *Gallus gallus* [GenBank: Q5ZIA8], *Tetraodon nigroviridis* [GenBank: CAF97838], *Oncorhynchus mykiss* [GenBank: AAZ66043], *Salmo salar* [GenBank: B5X4M8].

one helix and nine β-strands. Compared to the template, S13 is missing and S14 is shorter, both these reasons contributing to the existence of a large loop connecting S12 to S14. Finally, the Ig C-like domain contains seven β-strands in the model, with S20, S25 and H5 missing when compared to the template. In spite of the observed differences in SSEs, overall the predicted three-dimensional structure of sea bass tapasin matches that of the template.

Phylogenetic comparison of TPN protein sequences from several organisms reveals that among teleosts sea bass TPN is more closely related to that from *O. latipes* and *Oryzias dancena* (Fig. 6 and Table 2). Analysing the evolutionary relationship of TPN molecules through their Ig domains and those of other proteins (TPN-R, CD8α, β2m, MHC, and IgM light chain) clustered all V and C domains independently (Fig. 7), which is in agreement with Jorgensen et al. [37] results and broadens their suggestion that both TPN Ig domains are true V and C domains to more teleost tapasins. All TPN Ig-V domains cluster together with high bootstrapping segregating them from TPN-R Ig-V ones. The TPN Ig-C domains are not in a single cluster but instead appear mixed with TPN-R Ig-C domains, although low bootstrap levels are not supportive of such branching order.

Table 2
TPN amino acid percentages of similarity and identity determined with MatGat.

	Dicentrarchus labrax [GenBank: HQ328073]	
	Similarity (%)	Identity (%)
Oryzias latipes [GenBank: BAB83851]	88.8	76.4
Oryzias dancena [GenBank: ACN49160]	88.2	75.8
Oncorhynchus mykiss [GenBank: AAZ66039]	84.2	73.6
Salmo salar [GenBank: AB227285]	84.4	72.7
Danio rerio [GenBank: NP_571049]	83.5	67.5
Gallus gallus [GenBank: O73895]	50.2	32.5
Rattus norvegicus [GenBank: CAC34730]	49.6	31.1
Mus musculus [GenBank: AAD17964]	50.1	31.4
Homo sapiens [GenBank: AAC20076]	50.0	29.8

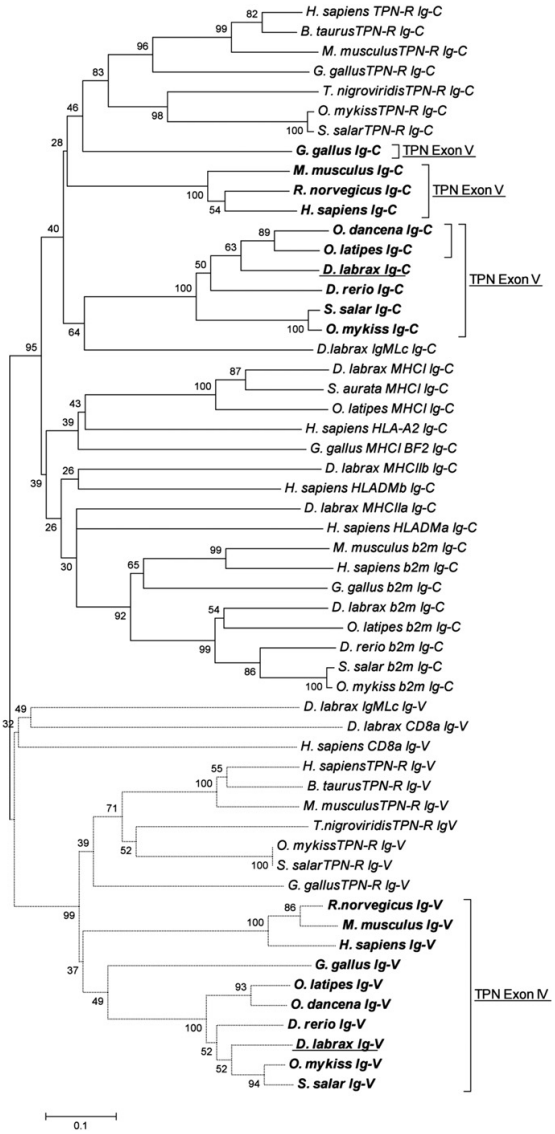


Fig. 7. Phylogenetic relationship between Ig domains of TPN, TPN-R, CD8α, β2m, MHC, and IgM generated with Mega version 3.1 [45]. The branches were validated by bootstrap analysis from 1000 replications, which are represented by percentages in branch nodes. Accession numbers from TPN sequences have been referred in figure 4, and those of TPN-R in figure 6. The remaining are the following: *H. sapiens* β2m IgC [GenBank: P61769], *D. labrax* β2m IgC [GenBank: HQ290128]; *H. sapiens* HLA-A2 IgC [GenBank: P01892], *G. gallus* MHCI BF2 IgC [GenBank: Q95601], *D. labrax* MHCI IgC [GenBank: HQ290107]; *D. labrax* MHCIIa IgC [GenBank: ABH09449], *D. labrax* MHCIIb IgC [GenBank: ABH09453], *H. sapiens* HLA-DMa [GenBank: NP_006111], *H. sapiens* HLA-DMb [GenBank: NP_002109]; *D. labrax* IgMLc IgV [GenBank: CAC16856], *D. labrax* IgMLc IgC [GenBank: CAC16856]; *D. labrax* CD8a IgV [GenBank: AAZ66439], *H. sapiens* CD8a IgV [GenBank: NP_741969].

In conclusion, in the present study the sea bass tapasin cDNA and gene were for the first time successfully cloned. Dila-TPN consists of a signal peptide, unique domain, IgSf V-like domain, IgSf C-like domain, transmembrane domain, and cytoplasmic tail

matching other known tapasins. The presented gene is homologous to the human one and characteristics like exon/intron boundaries, and the presence of a single N-glycosylation site and an ER retention motif, also resemble those from other tapasin genes. Although there are regions of insertions and deletions, within the N-domain and the IgSf-V domain of sea bass TPN when compared to the human molecule, the theoretical model of Dila-TPN is similar to the L-shaped structure of hTPN. Monoclonal and/or polyclonal antibodies against sea bass tapasin are currently not available. Thus, the present identification of sea bass tapasin provides information that will enable the production of new tools for studying the class I antigen presentation pathway in this important fish species.

Acknowledgements

Funding from FCT – Fundação para a Ciência e a Tecnologia, Portugal (POCTI/CVT/44925/2006), supported this work. Rute D. Pinto acknowledges FCT PhD fellowship BD/42327/2007, financed by POPH-QREN and co-funded by FSE and MCTES. The funders had no role in study design, data collection and analysis, decision to publish, or publication of the manuscript. No additional external funding was received for this study.

Appendix. Supplementary data

Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.fsi.2011.10.029.

References

[1] Pamer E, Cresswell P. Mechanisms of MHC class I–restricted antigen processing. Annu Rev Immunol 1998;16:323–58.
[2] Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. Nature 1987;329:506–12.
[3] Peaper DR, Cresswell P. Regulation of MHC class I assembly and peptide binding. Annu Rev Cell Dev Biol. 2008;24:343–68.
[4] Kim Y, Kang K, Kim I, Lee YJ, Oh C, Ryoo J, et al. Molecular mechanisms of MHC class I-antigen processing: redox considerations. Antioxid Redox Signal 2009; 11:907–36.
[5] Finley D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annu Rev Biochem 2009;78:477–513.
[6] Procko E, Gaudet R. Antigen processing and presentation: TAPping into ABC transporters. Curr Opin Immunol 2009;21:84–91.
[7] Blanchard N, Shastri N. Coping with loss of perfection in the MHC class I peptide repertoire. Curr Opin Immunol 2008;20:82–8.
[8] Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a "molecular ruler" mechanism. Proc Natl Acad Sci USA 2005;102:17107–12.
[9] Zhang Y, Baig E, Williams DB. Functions of ERp57 in the folding and assembly of major histocompatibility complex class I molecules. J Biol Chem 2006;281: 14622–31.
[10] Diedrich G, Bangia N, Pan M, Cresswell P. A role for calnexin in the assembly of the MHC class I loading complex in the endoplasmic reticulum. J Immunol 2001;166:1703–9.
[11] Wearsch PA, Cresswell P. The quality control of MHC class I peptide loading. Curr Opin Cell Biol 2008;20:624–31.
[12] Momburg F, Tan P. Tapasin-the keystone of the loading complex optimizing peptide binding by MHC class I molecules in the endoplasmic reticulum. Mol Immunol 2002;39:217–33.
[13] Sadasivan B, Lehner PJ, Ortmann B, Spies T, Cresswell P. Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. Immunity 1996;5:103–14.
[14] Ortmann B, Copeman J, Lehner PJ, Sadasivan B, Herberg JA, Grandea AG, et al. A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. Science 1997;277:1306–9.
[15] Chen M, Bouvier M. Analysis of interactions in a tapasin/class I complex provides a mechanism for peptide selection. EMBO J 2007;26:1681–90.
[16] Bangia N, Lehner PJ, Hughes EA, Surman M, Cresswell P. The N-terminal region of tapasin is required to stabilize the MHC class I loading complex. Eur J Immunol 1999;29:1858–70.
[17] Tan P, Kropshofer H, Mandelboim O, Bulbuc N, Hammerling GJ, Momburg F. Recruitment of MHC class I molecules by tapasin into the transporter associated with antigen processing-associated complex is essential for optimal peptide loading. J Immunol 2002;168:1950–60.
[18] Garbi N, Tiwari N, Momburg F, Hammerling GJ. A major role for tapasin as a stabilizer of the TAP peptide transporter and consequences for MHC class I expression. Eur J Immunol 2003;33:264–73.
[19] Praveen PV, Yaneva R, Kalbacher H, Springer S. Tapasin edits peptides on MHC class I molecules by accelerating peptide exchange. Eur J Immunol 2010;40:214–24.
[20] Peaper DR, Wearsch PA, Cresswell P. Tapasin and ERp57 form a stable disulfide-linked dimer within the MHC class I peptide-loading complex. EMBO J 2005;24:3613–23.
[21] Kienast A, Preuss M, Winkler M, Dick TP. Redox regulation of peptide receptivity of major histocompatibility complex class I molecules by ERp57 and tapasin. Nat Immunol 2007;8:864–72.
[22] Wearsch PA, Cresswell P. Selective loading of high-affinity peptides onto major histocompatibility complex class I molecules by the tapasin-ERp57 heterodimer. Nat Immunol 2007;8:873–81.
[23] Dong G, Wearsch PA, Peaper DR, Cresswell P, Reinisch KM. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase heterodimer. Immunity 2009;30:21–32.
[24] Frickel EM, Riek R, Jelesarov I, Helenius A, Wuthrich K, Ellgaard L. TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. Proc Natl Acad Sci USA 2002;99:1954–9.
[25] Wearsch PA, Jakob CA, Vallin A, Dwek RA, Rudd PM, Cresswell P. Major histocompatibility complex class I molecules expressed with monoglucosylated N-linked glycans bind calreticulin independently of their assembly status. J Biol Chem 2004;279:25112–21.
[26] Park B, Lee S, Kim E, Cho K, Riddell SR, Cho S, et al. Redox regulation facilitates optimal peptide selection by MHC class I during antigen processing. Cell 2006;127:369–82.
[27] Santos SG, Campbell EC, Lynch S, Wong V, Antoniou AN, Powis SJ. Major histocompatibility complex class I-ERp57-tapasin interactions within the peptide-loading complex. J Biol Chem 2007;282:17587–93.
[28] Herberg JA, Sgouros J, Jones T, Copeman J, Humphray SJ, Sheer D, et al. Genomic analysis of the Tapasin gene, located close to the TAP loci in the MHC. Eur J Immunol 1998;28:459–67.
[29] Grandea III AG, Comber PG, Wenderfer SE, Schoenhals G, Fruh K, Monaco JJ, et al. Sequence, linkage to H2-K, and function of mouse tapasin in MHC class I assembly. Immunogenetics 1998;48:260–5.
[30] Frangoulis B, Park I, Guillemot F, Severac V, Auffray C, Zoorob R. Identification of the Tapasin gene in the chicken major histocompatibility complex. Immunogenetics 1999;49:328–37.
[31] Clark MS, Shaw L, Kelly A, Snell P, Elgar G. Characterization of the MHC class I region of the Japanese pufferfish (*Fugu rubripes*). Immunogenetics 2001;52:174–85.
[32] Lukacs MF, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, et al. Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*). BMC Genomics 2007;8:251.
[33] Mehta RB, Nonaka MI, Nonaka M. Comparative genomic analysis of the major histocompatibility complex class I region in the teleost genus *Oryzias*. Immunogenetics 2009;61:385–99.
[34] Matsuo MY, Asakawa S, Shimizu N, Kimura H, Nonaka M. Nucleotide sequence of the MHC class I genomic region of a teleost, the medaka (*Oryzias latipes*). Immunogenetics 2002;53:930–40.
[35] Herrmann F, Trowsdale J, Huber C, Seliger B. Cloning and functional analyses of the mouse tapasin promoter. Immunogenetics 2003;55:379–88.
[36] Landis ED, Palti Y, Dekoning J, Drew R, Phillips RB, Hansen JD. Identification and regulatory analysis of rainbow trout tapasin and tapasin-related genes. Immunogenetics 2006;58:56–69.
[37] Jorgensen SM, Grimholt U, Gjoen T. Cloning and expression analysis of an Atlantic salmon (*Salmo salar* L.) tapasin gene. Dev Comp Immunol 2007;31: 708–19.
[38] Sultmann H, Murray BW, Klein J. Identification of seven genes in the major histocompatibility complex class I region of the zebrafish. Scand J Immunol 2000;51:577–85.
[39] Stet RJ, van Erp SH, Hermsen T, Sultmann HA, Egberts E. Polymorphism and estimation of the number of MhcCyca class I and class II genes in laboratory strains of the common carp (*Cyprinus carpio* L.). Dev Comp Immunol 1993;17:141–56.
[40] Higgins DG. CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol Biol. 1994;25:307–18.
[41] Hall T. BioEdit. Biological sequence alignment editor for Windows. NC, USA: North Carolina State University; 1998.
[42] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol. 2004;340:783–95.
[43] Zdobnov EM, Apweiler R. InterProScan – an integration platform for the signature-recognition methods in InterPro. Bioinformatics 2001;17:847–8.
[44] Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct Biol. 2009;9:51.
[45] Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 2004; 5:150–63.
[46] Campanella JJ, Bitincka L, Smalley J. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics 2003;4:29.
[47] Ausubel FM, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. Current protocols in molecular biology. New York: John Wiley and Sons, Inc; 1999.
[48] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195–201.

[49] Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 2010;66:213–21.
[50] Hutchinson EG, Thornton JM. PROMOTIF — a program to identify and analyze structural motifs in proteins. Protein Sci 1996;5:212–20.
[51] Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272.
[52] Laskowski RAMMW, Moss D, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–91.
[53] Williams AF, Barclay AN. The immunoglobulin superfamily-domains for cell surface recognition. Annu Rev Immunol 1988;6:381–405.
[54] Deverson EV, Powis SJ, Morrice NA, Herberg JA, Trowsdale J, Butcher GW. Rat tapasin: cDNA cloning and identification as a component of the class I MHC assembly complex. Genes Immun 2001;2:48–51.
[55] Dick TP, Bangia N, Peaper DR, Cresswell P. Disulfide bond isomerization and the assembly of MHC class I–peptide complexes. Immunity 2002;16:87–98.
[56] Simone LC, Wang X, Tuli A, Solheim JC. Effect of a tapasin mutant on the assembly of the mouse MHC class I molecule H2-K(d). Immunol Cell Biol 2010;88:57–62.
[57] Ioerger TR, Du C, Linthicum DS. Conservation of cys-cys trp structural triads and their geometry in the protein domains of immunoglobulin superfamily members. Mol Immunol 1999;36:373–86.
[58] Petersen JL, Hickman-Miller HD, McIlhaney MM, Vargas SE, Purcell AW, Hildebrand WH, et al. A charged amino acid residue in the transmembrane/cytoplasmic region of tapasin influences MHC class I assembly and maturation. J Immunol 2005;174:962–9.
[59] Papadopoulos M, Momburg F. Multiple residues in the transmembrane helix and connecting peptide of mouse tapasin stabilize the transporter associated with the antigen-processing TAP2 subunit. J Biol Chem 2007;282:9401–10.
[60] Simone LC, Wang X, Tuli A, McIlhaney MM, Solheim JC. Influence of the tapasin C terminus on the assembly of MHC class I allotypes. Immunogenetics 2009;61:43–54.
[61] Jackson MR, Nilsson T, Peterson PA. Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. EMBO J 1990;9:3153–62.
[62] Paulsson KM, Jevon M, Wang JW, Li S, Wang P. The double lysine motif of tapasin is a retrieval signal for retention of unstable MHC class I molecules in the endoplasmic reticulum. J Immunol 2006;176:7482–8.

Supplementary data

Table S1

Differences in Dila-TPN*01 compared to Dila-TPN*02 (excluding microsatellite).

	Feature	Replacements
Nucleotide	**Full Length**	18
	5' UTR	0
	Exon 1 (coding)	0
	Exon 2	3[a]
	Exon 3	3
	Exon 4	1[b]
	Exon 5	2[c]
	Exon 6	0
	Exon 7	0
	Exon 8 (coding)	0
	3' UTR	9
Protein	**Mature**	3

(a) 1 replacement translated in 1 different amino acid: A^{47}/T^{47} (numbering of the mature protein)

(b) 1 replacements translated in 1 different amino acids: R^{238}/H^{238} (numbering of the mature protein)

(c) 1 replacement translated in 1 different amino acid: K^{342}/T^{342} (numbering of the mature protein)

Table S2

Differences in Dila-TPN*01 compared to Dila-TPN*03 (excluding microsatellite).

	Feature	Replacements
Nucleotide	**Full Length**	17
	5' UTR	0
	Exon 1 (coding)	0
	Exon 2	2[a]
	Exon 3	2
	Exon 4	2[b]
	Exon 5	2[c]
	Exon 6	0
	Exon 7	0
	Exon 8 (coding)	0
	3'UTR	9
Protein	**Mature**	3

(a) 1 replacement translated in 1 different amino acid: A^{47}/T^{47} (numbering of the mature protein)

(b) 1 replacements translated in 1 different amino acids: S^{178}/L^{178} (numbering of the mature protein)

(c) 1 replacement translated in 1 different amino acid: K^{342}/T^{342} (numbering of the mature protein)

Table S3

Differences in Dila-TPN*02 compared to DilaTPN*03.

	Feature	Replacements	Insertions	Deletions
Gene	**Full Length**	67	18	18
	5' UTR	0	0	0
	Exon 1 (coding)	0	0	0
	Intron 1	22	6	3
	Exon 2	1	0	0
	Intron 2	2	0	0
	Exon 3	1	0	0
	Intron 3	3	0	0
	Exon 4	3[a]	0	0
	Intron 4	7	0	0
	Exon 5	2[b]	0	0
	Intron 5	10	0	0
	Exon 6	0	0	0
	Intron 6	4	2	15
	Exon 7	0	0	0
	Intron 7	0	0	0
	Exon 8 (coding)	0	0	0
	3'UTR	12	10	0
Protein	**Mature**	3	0	0

(a) 2 replacements translated in 2 different amino acids: L^{178}/S^{178} and R^{238}/H^{238} (numbering of the mature protein)

(b) 1 replacement translated in 1 different amino acid: K^{342}/T^{342} (numbering of the mature protein)

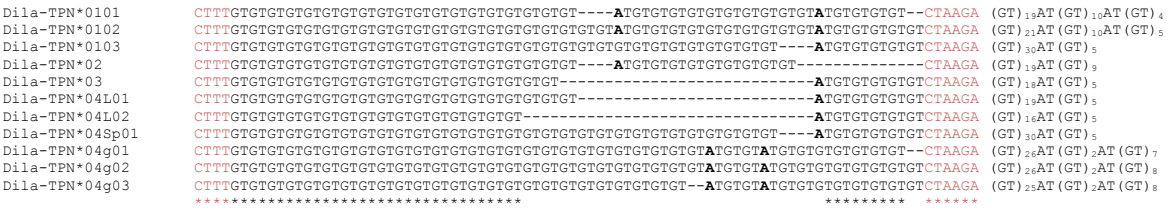


Figure S1. Microsatellite region of TPN transcripts and genes. Number of repeats are indicated for each molecule: $(GT)_n[AT]_1(GT)_n$ or $(GT)n[AT]_1(GT)n[AT]_1(GT)_n$. In order to overcome the difficulty to ascertain whether the differences observed between Dila-TPN*0101, *0102, *0103, *02 and *03 were due to technical error or could be occurring *in vivo*, a slightly broader region that included the microsatellite was assessed in a new individual so that cDNA (liver and spleen) and genomic DNA (gDNA) from the same fish could be used as templates. Primers DLTPNFW8 and DLTPNRV8 (Table 1) were used in the PCR reactions, and amplification products were purified, cloned and sequenced. Sequencing of three clones from each different template resulted in the here-represented additional sequences: two sequences from liver cDNA (Dila-TPN*04L01 and *04L02), one from spleen cDNA (Dila-TPN*04Sp01) and three from gDNA (Dila-TPN*04g01, *04g02 and *04g03). These results show that even using cDNA and gDNA from the same fish, differences were still observed in the microsatellite although with gaps in only one region. So, especially considering that TPN is a single-copy gene (for further details see Results and Discussion section) it is likely that the differences observed are due to PCR/sequencing errors.

CHAPTER 5

Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) calreticulin

Rute D. Pinto, Ana R. Moreira, Pedro J. B. Pereira and Nuno M. S. dos Santos

Fish & Shellfish Immunol. 2013, *in press*, doi: 10.1016/j.fsi.2013.03.004

Fish & Shellfish Immunology xxx (2013) 1–8

Contents lists available at SciVerse ScienceDirect

Fish & Shellfish Immunology

journal homepage: www.elsevier.com/locate/fsi

Short communication

Molecular cloning and characterization of sea bass (*Dicentrarchus labrax*, L.) calreticulin

Rute D. Pinto [a,c], Ana R. Moreira [a], Pedro J.B. Pereira [b], Nuno M.S. dos Santos [a,*]

[a] Fish Immunology and Vaccinology Group, Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal
[b] Biomolecular Structure Group, Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Rua do Campo Alegre 823, 4150-180 Porto, Portugal
[c] Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Rua de Jorge Viterbo Ferreira 228, 4050-313 Porto, Portugal

ARTICLE INFO

Article history:
Received 7 November 2012
Received in revised form
15 February 2013
Accepted 4 March 2013
Available online xxx

Keywords:
Dicentrarchus labrax
Calreticulin
Comparative sequence analysis
Protein structure prediction
Homology modelling

ABSTRACT

Mammalian calreticulin (CRT) is a key molecular chaperone and regulator of Ca^{2+} homeostasis in endoplasmic reticulum (ER), also being implicated in a variety of physiological/pathological processes outside the ER. Importantly, it is involved in assembly of MHC class I molecules. In this work, sea bass (*Dicentrarchus labrax*) CRT (Dila-CRT) gene and cDNA have been isolated and characterized. The mature protein retains two conserved motifs, three structural/functional domains (N, P and C), three type 1 and 2 motifs repeated in tandem, a conserved pair of cysteines and ER-retention motif. It is a single-copy gene composed of 9 exons. Dila-CRT three-dimensional homology models are consistent with the structural features described for mammalian molecules. Together, these results are supportive of a highly conserved structure of CRT through evolution. Moreover, the present data provides information that will allow further studies on sea bass CRT involvement in immunity and in particular class I antigen presentation.

1. Introduction

Calreticulin (CALR or CRT) was first identified as a high-affinity Ca^{2+}-binding protein in the sarcoplasmic reticulum of skeletal muscle [1]. Different designations have been attributed to it, but the accepted term calreticulin reflects both its calcium binding nature and its localization in the sarcoplasmic/endoplasmic reticulum (SR/ER) of muscular/non-muscular tissues, respectively (reviewed in [2]). Being present in vertebrates, invertebrates and higher plants (reviewed in [3]), its importance has been demonstrated by multiple functions both inside (chaperoning and Ca^{2+} storage/release) and outside the ER (many physiological/pathological processes) (reviewed in [4,5]).

Calreticulin can be divided into three domains (reviewed in [6]): the amino-terminal N domain, the flexible mid proline-rich P domain, both taking part in chaperone activity, and the highly acidic carboxyl-terminal C domain responsible for the Ca^{2+}-buffering activity (reviewed in [5,6]) that ends with the ER-retrieval sequence [7]. The overall shape of CRT includes a globular structure (N/C domains) [8–10] and an extended arm-like hairpin structure (P domain) [11]. Besides Ca^{2+} [12–14], CRT also binds ATP [15] and Zn^{2+} [16], which affect its ability to suppress aggregation of non-glycosylated proteins [17]. Hence, CRT acts as a lectin chaperone (interacts with monoglycosylated proteins) and as a classical chaperone (has polypeptide-binding capacity/senses a protein conformational state) [17].

Of notice is the involvement of CRT in major histocompatibility complex (MHC) class I folding and high affinity peptide loading (reviewed in [18,19]). Importantly, assembly occurs within the ER (reviewed in [20,21]) and is dependent on CRT [22].

Studies of calreticulin in teleost fish are still limited to a few species [23–27]. Here we report the identification and characterization of calreticulin from European sea bass, a relevant species for the Mediterranean aquaculture industry. Full cDNA and gene sequences as well as gene copy number have been determined. Moreover, *in silico* structural (primary sequence and 3D homology modelling) and phylogenetic analyses have been performed in order to further characterize this molecule.

2. Materials and methods

2.1. Fish

Sea bass, *D. labrax*, were kept in a recirculating, ozone-treated salt-water (20–25‰) system at 22 ± 1 °C and fed with commercial pellets twice a day. Fish were sacrificed with a lethal dose of 2-phenoxyethanol (Panreac; >5 mL/10 L).

* Corresponding author. Tel.: +351 226 074 900; fax: +351 226 099 157.
E-mail addresses: rsp@ibmc.up.pt (R.D. Pinto), amoreira@ibmc.up.pt (A.R. Moreira), ppereira@ibmc.up.pt (P.J.B. Pereira), nsantos@ibmc.up.pt, nmsdossantos@hotmail.com (N.M.S. dos Santos).

1050-4648/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsi.2013.03.004

Fig. 1. Alignment of CRT amino acid sequences. Amino acid sequences were retrieved from GenBank (Table S2) and aligned with CLUSTALW [50]. Dashes indicate gaps that maximize the alignment and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with, (*), (:) and (.), respectively. Residue numbers are given right of the sequences; numbers in parenthesis refer to the mature proteins. Predicted N, P and C domains are indicated above the alignment [10], with junction regions represented by black boxes (open ones along the alignment). Letters (s) or (h) above the alignment denote β-strands or α-helices, respectively; different β-sheets within the globular domain are denoted with different colours. The two CRT signature motifs within the N domain are boxed and labelled (Calreticulin-1 and -2). Type 1 (or A) and 2 (or B) tandem repeats are indicated above the alignment (also underlined and in bold type in the sea bass sequence). Putative nuclear localization signal is shaded in grey. The region interacting with rheumatoid arthritis shared epitope is shaded cyan in the rabbit sequence [46]. Residues of human CRT that contact ERp57 [43] are shaded green. Canonical cysteines involved in an intra-chain disulfide bond (connecting β6 and β7) as well as a third cysteine (absent in *Rana rugosa*) are displayed in inverted type. Histidines within the N domain (putative sites for Zn^{2+} binding) are shaded in yellow. Mouse residues that can putatively bind Zn^{2+} are shaded orange [9]. Putative N-glycosylation signal identified in some species is shaded dark blue. Mouse and human CRT residues known to interact with Ca^{2+} in the globular domain are shaded dark red [8,9]. Residues involved in carbohydrate binding are shaded light (hydrogen bonds) and dark (hydrophobic interactions) purple [9]. Hydrophobic patch on the P domain putatively involved in peptide binding is shaded red [10]. Residues involved in the formation of three small hydrophobic cores in the rat P domain are underlined and in bold type [11]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.2. cDNA cloning

RNAs were extracted according to Versagene™ RNA tissue kit (Gentra Systems) instructions and transcribed to cDNAs following the BioScript RNase H^- (Bioline) protocol. PCR products were cloned into pGEM®-T Easy (Promega). Automated sequencing was performed using primers detailed in Supplementary Table S1. 5′ RACE experiments were performed with the 5′ RACE System from Invitrogen (Version 2.0) and Recombinant Terminal Transferase (Fermentas or Roche) was used to dATP tail the purified cDNA.

Total RNA extracted from the head-kidney of one fish was reverse transcribed with primer APv2 (Table S1). Degenerate primers were designed based on conserved regions of calreticulin across several vertebrate species. The cDNA was amplified with primers CnxCrtFW1 and CnxCrtRV1 (Table S1) and products were obtained at the 4th round of PCR. A product of approximately 400 bp was purified, cloned and sequenced.

This first product from sea bass calreticulin was used to design specific reverse primers. As the distance to the 5′ end of the transcript was still considerable, also a degenerate forward primer was designed using the calnexin/calreticulin multiple alignment. Total RNA from head-kidney was synthesized with APv2 and amplified with the set of primers CrtFw1/DLCrtRV5 followed by two consecutive amplifications with CrtFW1/DLCrtRV4 (Table S1). The obtained ~350 bp product was cloned and sequenced.

To obtain the 5′ UTR of calreticulin, the resulting sequence was used to design two specific reverse primers (Table S1). Total RNA extracted from the spleen of one fish was synthesized with DLCrtRV7. The tailed cDNA was amplified with primers APv2/DLCrtRV7 followed by a 2nd amplification with AUAP2/DLCrtRV6 (Table S1). The obtained ~400 bp product was purified, cloned and sequenced.

To obtain the full sea bass calreticulin cDNA, specific primers (Table S1) were designed at the beginning of the 5′ UTR region.

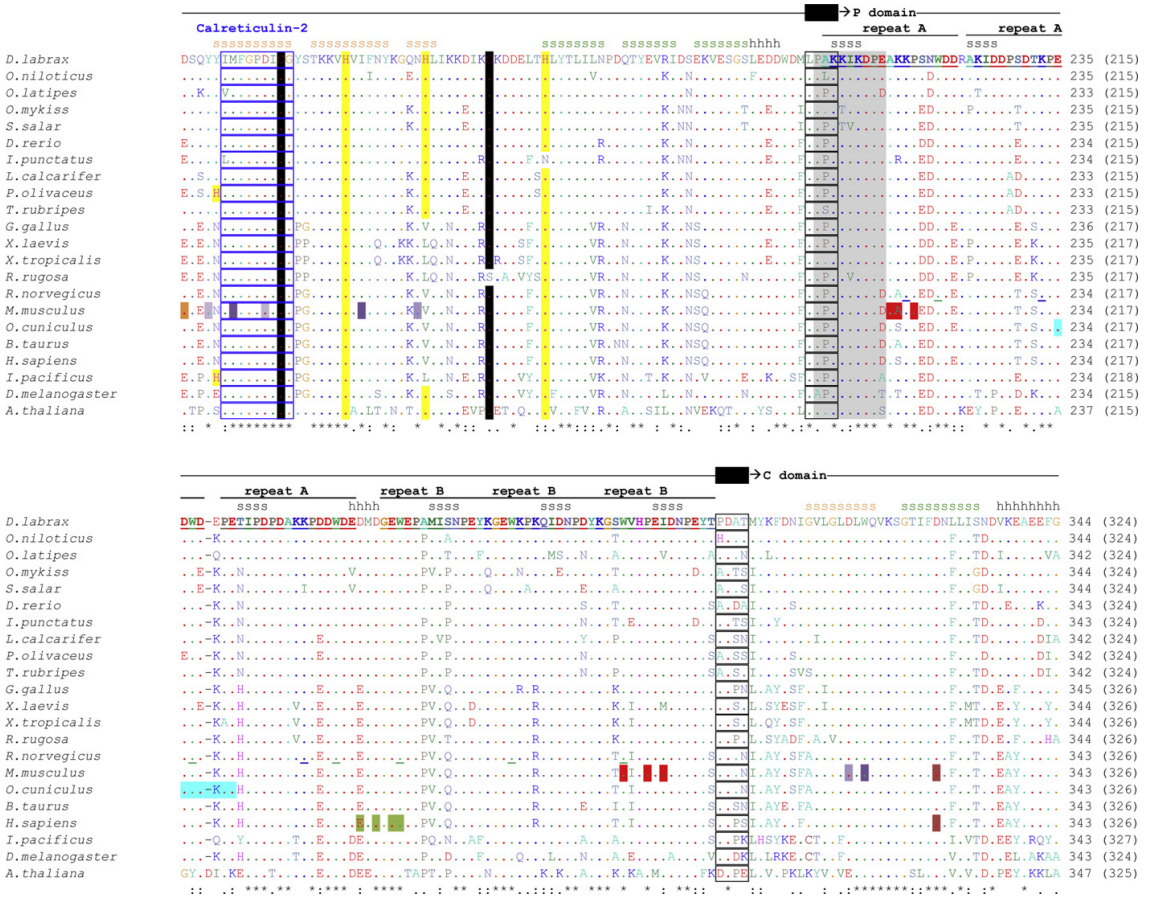


Fig. 1. (*continued*).

DLCrtFW1 and APV were used in a first amplification using liver cDNA, followed by a semi-nested PCR with primers DLCrtFW2/AUAP. The obtained ~2100 bp PCR product was purified, cloned and 3 independent clones were sequenced.

2.3. Genomic DNA cloning and Southern blotting

Genomic DNA (gDNA) was isolated from sea bass erythrocytes from a single fish [28]. The calreticulin gene was obtained by partial amplification using different sets of primers (Table S1). PCR products were purified, cloned and one clone of each product sequenced as before. Full calreticulin gene was assembled by overlapping the partial sequences.

Sea bass genomic DNA was digested with different restriction enzymes and subjected to Southern blotting [32], as detailed in Fig. S3.

2.4. Sequence analysis

Nucleotide and protein sequences from sea bass calreticulin were compared to several homologues currently available at the GenBank database [http://www.ncbi.nlm.nih.gov/genbank/]. The multiple sequence alignment was made using CLUSTALW [29] and formatted with Bioedit [30]. Domain boundaries were derived from the murine structure [10]. Calreticulin family signature motifs were based on InterPro Scan predictions [31]. The molecular weights of the polypeptide chains were calculated with the Expasy compute pI/MW tool [http://www.expasy.org]. N-glycosylation sites and putative signal peptide sequences were predicted at [http://www.cbs.dtu.dk/services/NetNGlyc/] and [http://www.cbs.dtu.dk/services/SignalP/], respectively.

2.5. 3D modelling

Suitable structural templates for sea bass calreticulin, mouse (PDB: 3rg0) and rat (PDB: 1hhn) calreticulins, were identified by a BLAST search as implemented in the SWISS-MODEL Protein Modelling Server [33]. Homology modelling was performed with SWISS-MODEL in automated mode [33] revealing ~62% and ~79% identity between target and template molecules, respectively. Gaps were introduced only in the truncated area of the template 3rg0, in this case involving secondary structure elements. The secondary structure elements were compared with PROMOTIF [34]. Model quality was assessed with WHATCHECK [35] and PROCHECK [36] as implemented in the Swiss Model structure assessment tool. Model pictures were prepared with PyMOL [http://pymol.org/].

Fig. 1. (*continued*).

3. Results and discussion

3.1. cDNA, gene and Southern blot analysis

The sea bass calreticulin (Dila-CRT) full-length cDNA has been obtained by homology cloning, performing RT-PCR on mRNA extracted from the liver of one fish. Sequence analysis confirmed highest similarity with other fish CRTs. The here-reported Dila-CRT sequence [GenBank: JX235975] has a total size of 2065 bp, containing a 5′ UTR of 34 bp, an ORF of 1281 bp and a 3′ UTR of 750 bp (excluding the poly(A) tail). Three mRNA instability motifs [37] and two potential polyadenylation signals could be identified within the 3′ UTR.

In the Dila-CRT gene [Genbank: JX235974], which is 6462 bp long, all intron donor/acceptor sites follow the gt/ag consensus rule, and codons in the exon boundaries are split by phase-one (1–3), or phase-zero (4–8) introns (Fig. S1). Exon I encodes the 5′ UTR, the putative leader peptide and the first 34 nucleotides of mature CRT – the start of the N domain; exons II-IV cover most of the N domain length; exon V encodes the final part of N domain (last 99 nucleotides) and the beginning of the P domain (111 nucleotides), which continues through exon VI; exon VII encodes P domain final 105 nucleotides, as well as the first part of the C domain (39 nucleotides) that continues through exon VIII; finally exon IX contains the last part of the C domain and the 3′ UTR (Fig. S1). The sea bass CRT nine-exon genomic structure (Fig. S2) is identical to that of human [38], mouse [39] and fish CRTs [23,25,27]. Furthermore, although intron sizes are highly variable, intron phases and locations as well as exon sizes are conserved between sea bass, human [38] and other species [27] CRTs.

Southern blot analysis revealed that Dila-CRT is a single-copy gene (Fig. S3), which is in agreement to what has been described not only for human [40] and mouse [39] CRTs but also for its teleost homologues including zebrafish [23] and rainbow trout [25], despite rainbow trout tetraploidy.

3.2. Primary structure analysis

Amino acid sequences of CRT from several species (Table S2) were aligned with that of Dila-CRT using CLUSTALW (Fig. 1). In sea bass, a predicted 406-residue (47.2 kDa) mature protein would be generated by cleavage of the putative 20-amino acid signal peptide. It is the second largest CRT among all analysed species (Fig. 1).

Sea bass CRT contains the typical N, P and C CRT domains, the two first displaying higher identity/similarity among the different species than the C domains (Fig. 1). Within the N domain two conserved CRT family signature motifs (K^{99}HEQKIDCGGGYVKIF^{114} and I^{131}MFGPDICG^{139}) were identified in Dila-CRT (Fig. 1), both consisting of highly conserved regions. Each of these motifs bears a conserved cysteine residue (in sea bass -sb- C^{106}, C^{138}) involved in the formation of an intramolecular disulfide bridge [41]. The equivalent linkage is required for efficient carbohydrate binding to mouse CRT [9]. However, in bovine brain CRT, a third also highly conserved cysteine residue at the N-domain (sbC^{144}) was shown to be involved instead [42]. CRT binds Zn^{2+} [16] with low affinity [13], presumably through five highly conserved histidine residues located at its N-domain [16] (sbH^{43}, H^{100}, H^{124}, H^{146} and H^{171} in Fig. 1). Teleosts display an additional histidine within this domain (sbH^{156}), also observed in *Drosophila melanogaster* and *Arabidopsis thaliana* (Fig. 1). Although some of these residues are buried, mouse H^{42} is exposed and in the vicinity of other residues that could potentially participate in Zn^{2+} binding: D^{118}, D^{121}, H^{123} and D^{125} [9], all highly conserved across species (Fig. 1).

In the arm/P domain, the triplicate A (or type 1) and B or (type 2) repeats (PXXIXDPDAXKPEDWDE and GXWXPPIXNPXYX, respectively) ordered as '111222' as in other described calreticulins (reviewed in [3]) are also observed in Dila-CRT (Fig. 1): **A** – A^{206}…D^{222}, A^{224}…D^{238} and P^{240}…E^{256}; **B** – G^{260}…K^{273}, G^{274}…K^{287} and G^{288}…T^{301} (only three non-conservative replacements among all six motifs in sea bass – D^{231}, K^{279} and H^{292}). A putative nuclear localization signal (NLS) – P^{204}PKKIKDPD – identified in rabbit CRT P domain (reviewed in [2]) is highly conserved among the different species, with only two conservative substitutions in Dila-CRT: P^{205}AKKIKDPE. The tip of the P domain is the binding site to ERp57 [43], and mutational analysis on rabbit CRT identified some of the residues involved: E^{239}, D^{241}, E^{243} and W^{244} [44]. With exception of E^{239} that is conservatively substituted by an aspartate in teleosts (sbD^{237}), the remaining residues are conserved across species (sbD^{239}, sbE^{241}, sbW^{242}) (Fig. 1), suggesting that binding to ERp57 may be conserved as well. CRT has long been hypothesized

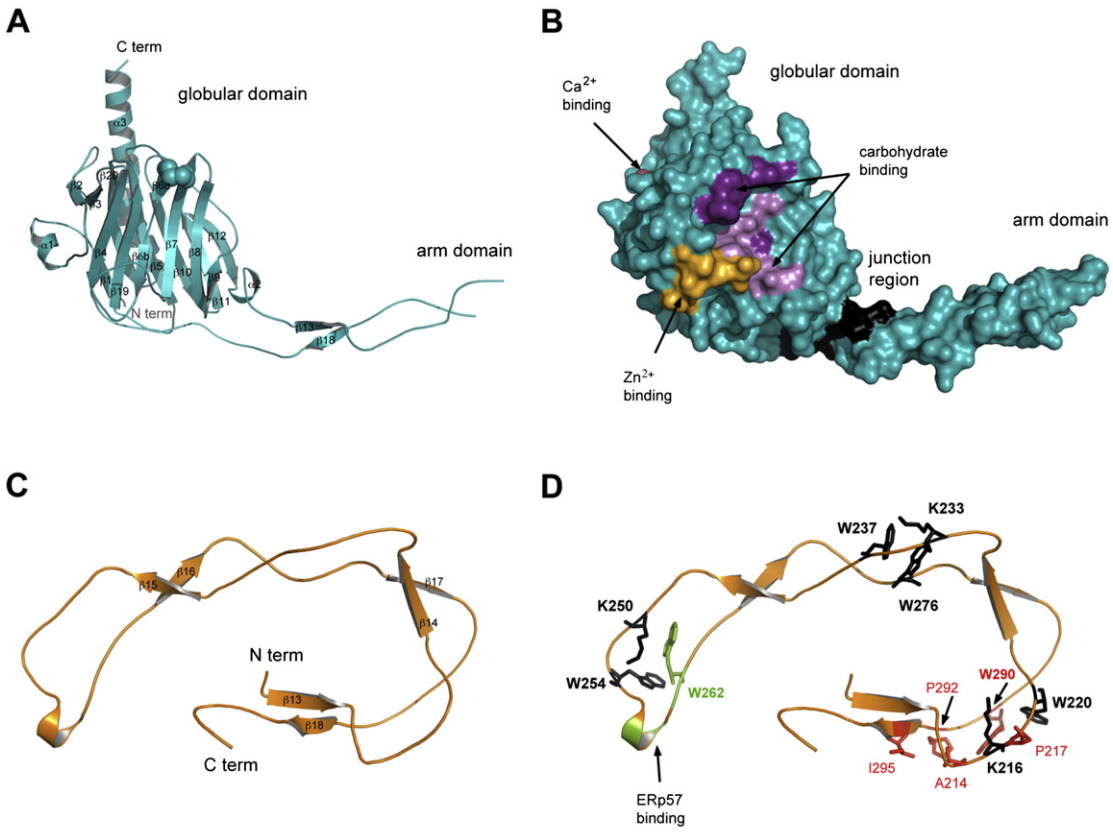


Fig. 2. Three-dimensional homology models of sea bass CRT globular/partial arm domains and arm domain alone. Models generated using the crystal structure of mouse CRT [PDB: 3rg0] [10] and rat CRT NMR structure [PDB: 1hhn] [11] as templates. **A) Dila-CRT globular/partial arm domain homology model**. Secondary structure elements (SSE), as well as N- and C-termini are labelled. The disulfide bond is represented as spheres. Note that C domain contributions to the globular region include: strands β19 and β20 and helix α3. **B) Surface representation of the Dila-CRT globular/partial arm domain homology model**. The Ca^{2+}-, glycan- and putative Zn^{2+}-binding sites are labelled and coloured dark red, purple (light for hydrogen bonds and dark for hydrophobic contacts) and orange, respectively. **C) Dila-CRT arm domain homology model**. The six β-sheet strands are labelled. N and C terminal regions are also indicated. **D) Same as panel C with relevant residues highlighted**. The ERp57-binding site is coloured green. Residues putatively involved in complementary peptide binding are coloured red and labelled. Residues forming the three hydrophobic cores of the P domain are coloured black and labelled. **E) Alignment of target and template amino acid sequences**. The protein sequences of CRT from *Dicentrarchus labrax*, *Mus musculus* and *Rattus norvegicus* were used in the multiple sequence alignment (see details in Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

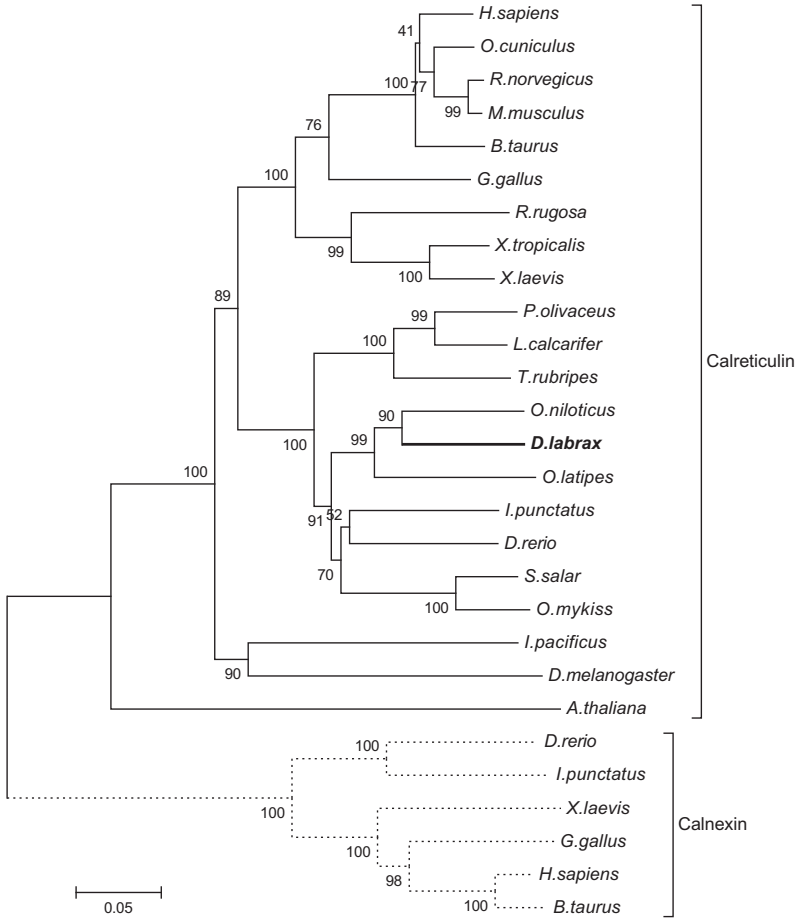


Fig. 3. Phylogenetic tree of CRT amino acid sequences from different vertebrate species. Neighbour-joining tree calculated with MEGA v5.05 [51], using p-distance parameter and pairwise deletion of gaps. The branches were validated by bootstrap analysis from 1000 replications, represented by percentages in branch nodes. The sequences of precursor forms of CRT were used. Calnexin molecules from distinct species were used as outgroup (accession numbers detailed in Table S2).

to play a role in autoimmunity [45]. Ling S. et al. [46] findings indicate that the rheumatoid arthritis shared epitope (SE – a five residue motif from HLA-DRB1 alleles known as a risk factor for disease severity [47] and to interact with cell surface CRT [48]) activated signaling maps to the 234–240 region of the (rabbit) CRT P-domain and depends primarily on E^{234} and E^{240}, and to a lesser extent on D^{237} [46]. These key residues are absolutely conserved in Dila-CRT (E^{235}, E^{241}, and D^{238}) and highly conserved across species (Fig. 1) suggesting functional relevance.

No potential glycosylation site has been predicted for Dila-CRT. Depending on the species, calreticulin may have none, one or more potential glycosylation sites, suggesting that the glycosylation pattern is not a conserved feature of the protein (reviewed in [3]). Similarly to other CRT molecules, Dila-CRT C domain is highly acidic, containing 38 glutamate and 15 aspartate residues in a total of 118 amino acids (32 and 12%, respectively) (Fig. 1). In mammals, these series of negatively charged residues at the carboxy-terminus of the protein bind over 50% of the ER luminal Ca^{2+} – buffering activity (reviewed in [5,6]). Similar highly charged C-terminal domains have been reported to be in part responsible for the anomalous SDS-PAGE mobility of mammalian (reviewed in [2]) and fish [24] CRT at ~60 kDa instead of the expected ~46 kDa. Finally, Dila-CRT ends with an ER-retrieval sequence (Fig. 1) similar to most CRTs across the analyzed species. Variants (HEEL and HDEL) to the canonical KDEL signal were found in *Ixodes pacificus, D. melanogaster* and *A. thaliana*. Absence of the typical motif in *Gallus gallus* CRT is most probably a consequence of the molecule being incomplete.

Both CNX and CRT lectin/globular domains contain a single high affinity Ca^{2+}-binding site [9,49]. As revealed by the crystal structures of human and mouse CRT globular domains, the bound Ca^{2+} ion, playing a crucial stabilization role, is coordinated by the following ligands: the backbone carbonyls of Q^{26}, K^{62}, K^{64}, two water molecules and the side-chain carbonyl oxygens of D^{328} [8,9]. The Ca^{2+} ion, bound on the opposite site of the arm domain insertion [10], does not interfere with glycan recognition as was elucidated by Kozlov et al. [9] and the carbohydrate-binding site, located at the concave β-sheet, correlates well with previous mutagenesis studies identifying CRT residues essential for carbohydrate interactions [9]. Namely, residues Y^{109}, K^{111}, Y^{128}, D^{135}, D^{317} and D^{125} are involved in polar contacts, while M^{131} and W^{319} are

engaged in hydrophobic interactions with carbohydrate moieties [9]. Other contacts include G^{124} and N^{154} (H bonds), F^{74}, C^{105}, C^{137}, H^{145} and I^{147} (hydrophobic) [9]. All these residues show strict conservation across species, with the exception of the conservative substitutions of aspartates 125 and 317 for glutamates in some species (Fig. 1). According to Pocanschi et al. [10], within the lectin domain there is also a polypeptide-binding site that mediates CRT classical chaperone activity, and which precise location remains unknown. A hydrophobic patch along the P domain possibly constitutes a secondary polypeptide-binding site [10]. Residues A^{213}, A^{214}, P^{216}, W^{289}, P^{292} and I^{294} were found to contact other CRT molecule, of which only A^{214} is not conserved (Fig. 1). Three small hydrophobic cores have been described within the P domain each involving two tryptophans and one lysine [11], all conserved in sea bass ($K^{216}/W^{220}/W^{290}$; $K^{233}/W^{237}/W^{276}$; $K^{250}/W^{254}/W^{262}$) and almost strictly conserved across species (Fig. 1).

The arm domain is linked to the lectin domain by a small structured region, the junction region [10] that includes L^{203}–K^{206} and P^{301}–N^{304} on each end, structured around the aromatic ring of Y^{299}, forming a small hydrophobic core [10]. All these residues are conserved or conservatively substituted in sea bass (Fig. 1).

3.3. 3D modelling

The crystal structure of the ER luminal region of the dog calnexin revealed two main structural components: a globular domain and an extended arm-like domain [49]. There is no experimental structure of any full-length CRT available, but the structures of CRT fragments have been determined: rat CRT arm domain [11], mouse CRT globular domain (with bound carbohydrate) [9], human CRT globular domain [8] and mouse CRT globular/partial arm domains [10]. All these turned out to be suitable templates for modelling the sea bass molecule. The Dila-CRT structure was predicted by comparative modelling using the mouse CRT [PDB: 3rg0 [10]] and rat CRT [PDB: 1hhn [11]] crystal and NMR structures as templates for the globular/partial arm domains and arm domain alone, respectively (Fig. 2). As expected, the homology model of Dila-CRT is structurally compatible with the globular/lectin and arm/P domains described for other CRT molecules (Fig. 2). As described for the crystal structures of human and mouse calreticulins [8–10] (and also the homologous CNX [49]), the modelled sea bass CRT globular domain assumes a β-sandwich structure assembled from a seven-stranded concave β-sheet and a six-stranded convex β-sheet (with a small β-sheet covering the space between the first two, two short α-helices and a long α-helix that runs along and beyond the convex β-sheet) (Fig. 2A). Importantly, as mentioned for the mammalian CRT structures [8–10], the C domain is part of the globular domain, providing the central strands β19 and β20 (Fig. 2A). Functionally and structurally relevant residues involved in Ca^{2+}-, carbohydrate- and Zn^{2+}-binding that are conserved in Dila-CRT were mapped to the 3rg0-based model (Fig. 2B and E). Also in accordance to rat CRT structural report [11], the modelled Dila-CRT arm domain consists of an extended hairpin fold, involving the entire P domain sequence with both chain ends in close proximity, stabilized by three anti-parallel β-sheets with a helical turn at the tip of the hairpin (Fig. 2C). Again, conserved residues involved in ERp57 and secondary peptide binding, as well as those forming the three hydrophobic cores were mapped to the 1hhn-based model (Fig. 2D and E).

3.4. Phylogenetic analysis

In order to analyse the evolutionary relationships of CRT proteins from different organisms, a neighbour-joining tree was constructed using calnexin as an out-group (Fig. 3), with both groups of proteins clearly separated. Within the calreticulin group, each class of organisms segregates with high bootstrapping. As expected, Dila-CRT clusters with other bony fish (Actinopterygii) sequences being closely related to the *Oreochromis niloticus* one, from the same order (Perciformes). Additionally, a series of pair-wise alignments were performed with MatGAT in order to calculate similarity and identity of CRT amino acid sequences between species. The results confirm those previously obtained with the neighbour-joining method (Table S3): highest identity is observed with *O. niloticus* CRT (~85%) and the lowest to the plant molecule (~52%). Considering the different classes of organisms analysed, those showing more similarity towards Dila-CRT are the bony fish followed by mammals, interestingly with human CRT showing highest identity (~71%) among the mammalian species.

4. Conclusions

In the present study, the CRT transcript and gene from European sea bass were successfully cloned. All characteristic domains and critical features present in CRTs from other species are also found in Dila-CRT. Furthermore, characteristics such as gene organization and copy-number are also conserved between species. Phylogenetic analysis places sea bass CRT among the Actinopterygii cluster, with a high degree of similarity between CRTs from different species. Dila-CRT 3D homology models have been compared to those of other vertebrate molecules. Altogether these results indicate that the reported molecule is orthologue of the other CRTs described for other species, providing important primary data for future studies on diverse aspects of CRT chaperoning function, namely the MHC class I antigen presentation pathway in this species.

Acknowledgements

FEDER funds through the Operational Competitiveness Programme – COMPETE and national funds through FCT – Fundação para a Ciência e a Tecnologia under projects FCOMP-01-0124-FEDER-022718 (PEst-C/SAU/LA0002/2011) and FCOMP-01-0124-FEDER-007173 (POCTI/CVT/44925/2006) supported this work. Rute D. Pinto acknowledges FCT PhD fellowship BD/42327/2007, financed by POPH-QREN and co-funded by FSE and MCTES. The funders had no role in study design, data collection and analysis, decision to publish, or publication of the manuscript. No additional external funding was received for this study.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.fsi.2013.03.004.

References

[1] Ostwald TJ, MacLennan DH. Isolation of a high affinity calcium-binding protein from sarcoplasmic reticulum. J Biol Chem 1974;249:974–9.
[2] Michalak M, Milner RE, Burns K, Opas M. Calreticulin. Biochem J 1992;285(Pt 3):681–92.
[3] Michalak M, Corbett EF, Mesaeli N, Nakamura K, Opas M. Calreticulin: one protein, one gene, many functions. Biochem J 1999;344(Pt 2):281–92.
[4] Gold LI, Eggleton P, Sweetwyne MT, Van Duyn LB, Greives MR, Naylor SM, et al. Calreticulin: non-endoplasmic reticulum functions in physiology and disease. FASEB J 2010;24:665–83.
[5] Wang WA, Groenendyk J, Michalak M. Calreticulin signaling in health and disease. Int J Biochem Cell Biol 2012;44:842–6.
[6] Michalak M, Groenendyk J, Szabo E, Gold LI, Opas M. Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. Biochem J 2009;417:651–66.
[7] Fliegel L, Burns K, MacLennan DH, Reithmeier RA, Michalak M. Molecular cloning of the high affinity calcium-binding protein (calreticulin) of skeletal muscle sarcoplasmic reticulum. J Biol Chem 1989;264:21522–8.

[8] Chouquet A, Paidassi H, Ling WL, Frachet P, Houen G, Arlaud GJ, et al. X-ray structure of the human calreticulin globular domain reveals a peptide-binding area and suggests a multi-molecular mechanism. PLoS One 2011;6:e17886.
[9] Kozlov G, Pocanschi CL, Rosenauer A, Bastos-Aristizabal S, Gorelik A, Williams DB, et al. Structural basis of carbohydrate recognition by calreticulin. J Biol Chem 2010;285:38612–20.
[10] Pocanschi CL, Kozlov G, Brockmeier U, Brockmeier A, Williams DB, Gehring K. Structural and functional relationships between the lectin and arm domains of calreticulin. J Biol Chem 2011;286:27266–77.
[11] Ellgaard L, Riek R, Herrmann T, Guntert P, Braun D, Helenius A, et al. NMR structure of the calreticulin P-domain. Proc Natl Acad Sci U S A 2001;98:3133–8.
[12] Baksh S, Michalak M. Expression of calreticulin in Escherichia coli and identification of its Ca2+ binding domains. J Biol Chem 1991;266:21458–65.
[13] Li Z, Stafford WF, Bouvier M. The metal ion binding properties of calreticulin modulate its conformational flexibility and thermal stability. Biochemistry 2001;40:11193–201.
[14] Tan Y, Chen M, Li Z, Mabuchi K, Bouvier M. The calcium- and zinc-responsive regions of calreticulin reside strictly in the N-/C-domain. Biochim Biophys Acta 2006;1760:745–53.
[15] Corbett EF, Michalak KM, Oikawa K, Johnson S, Campbell ID, Eggleton P, et al. The conformation of calreticulin is influenced by the endoplasmic reticulum luminal environment. J Biol Chem 2000;275:27177–85.
[16] Baksh S, Spamer C, Heilmann C, Michalak M. Identification of the Zn2+ binding region in calreticulin. FEBS Lett 1995;376:53–7.
[17] Saito Y, Ihara Y, Leach MR, Cohen-Doyle MF, Williams DB. Calreticulin functions in vitro as a molecular chaperone for both glycosylated and non-glycosylated proteins. Embo J 1999;18:6718–29.
[18] Peaper DR, Cresswell P. Regulation of MHC class I assembly and peptide binding. Annu Rev Cell Dev Biol 2008;24:343–68.
[19] Wearsch PA, Cresswell P. The quality control of MHC class I peptide loading. Curr Opin Cell Biol 2008;20:624–31.
[20] Scholz C, Tampe R. The peptide-loading complex–antigen translocation and MHC class I loading. Biol Chem 2009;390:783–94.
[21] Van Hateren A, James E, Bailey A, Phillips A, Dalchau N, Elliott T. The cell biology of major histocompatibility complex class I assembly: towards a molecular understanding. Tissue Antigens 2010;76:259–75.
[22] Gao B, Adhikari R, Howarth M, Nakamura K, Gold MC, Hill AB, et al. Assembly and antigen-presenting function of MHC class I molecules in cells lacking the ER chaperone calreticulin. Immunity 2002;16:99–109.
[23] Rubinstein AL, Lee D, Luo R, Henion PD, Halpern ME. Genes dependent on zebrafish cyclops function identified by AFLP differential gene expression screen. Genesis 2000;26:86–97.
[24] Kales SC, Bols NC, Dixon B. Calreticulin in rainbow trout: a limited response to endoplasmic reticulum (ER) stress. Comp Biochem Physiol B Biochem Mol Biol 2007;147:607–15.
[25] Kales S, Fujiki K, Dixon B. Molecular cloning and characterization of calreticulin from rainbow trout (Oncorhynchus mykiss). Immunogenetics 2004;55: 717–23.
[26] Leong JS, Jantzen SG, von Schalburg KR, Cooper GA, Messmer AM, Liao NY, et al. Salmo salar and Esox lucius full-length cDNA sequences reveal changes in evolutionary pressures on a post-tetraploidization genome. BMC Genomics 2010;11:279.
[27] Liu H, Peatman E, Wang W, Abernathy J, Liu S, Kucuktas H, et al. Molecular responses of calreticulin genes to iron overload and bacterial challenge in channel catfish (Ictalurus punctatus). Dev Comp Immunol 2011;35:267–72.
[28] Stet RJ, van Erp SH, Hermsen T, Sultmann HA, Egberts E. Polymorphism and estimation of the number of MhcCyca class I and class II genes in laboratory strains of the common carp (Cyprinus carpio L.). Dev Comp Immunol 1993;17: 141–56.
[29] Higgins DG. CLUSTAL V: multiple alignment of DNA and protein sequences. Methods Mol Biol 1994;25:307–18.
[30] Hall T. BioEdit. Biological sequence alignment editor for windows. NC, USA: North Carolina State University; 1998.
[31] Zdobnov EM, Apweiler R. InterProScan–an integration platform for the signature-recognition methods in InterPro. Bioinformatics 2001;17:847–8.
[32] Ausubel FM, B R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al. Protocols in molecular biology. New York: John Wiley and Sons, Inc; 1999.
[33] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195–201.
[34] Hutchinson EG, Thornton JM. PROMOTIF–a program to identify and analyze structural motifs in proteins. Protein Sci 1996;5:212–20.
[35] Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature 1996;381:272.
[36] Laskowski RA, M MW, Moss D, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst 1993;26:283–91.
[37] Shaw G, Kamen R. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell 1986;46: 659–67.
[38] McCauliffe DP, Yang YS, Wilson J, Sontheimer RD, Capra JD. The 5′-flanking region of the human calreticulin gene shares homology with the human GRP78, GRP94, and protein disulfide isomerase promoters. J Biol Chem 1992;267:2557–62.
[39] Krause KH, Michalak M. Calreticulin Cell 1997;88:439–43.
[40] McCauliffe DP, Lux FA, Lieu TS, Sanz I, Hanke J, Newkirk MM, et al. Molecular cloning, expression, and chromosome 19 localization of a human Ro/SS-A autoantigen. J Clin Invest 1990;85:1379–91.
[41] Hojrup P, Roepstorff P, Houen G. Human placental calreticulin characterization of domain structure and post-translational modifications. Eur J Biochem 2001;268:2558–65.
[42] Matsuoka K, Seta K, Yamakawa Y, Okuyama T, Shinoda T, Isobe T. Covalent structure of bovine brain calreticulin. Biochem J 1994;298(Pt 2):435–42.
[43] Frickel EM, Riek R, Jelesarov I, Helenius A, Wuthrich K, Ellgaard L. TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. Proc Natl Acad Sci U S A 2002;99:1954–9.
[44] Martin V, Groenendyk J, Steiner SS, Guo L, Dabrowska M, Parker JM, et al. Identification by mutational analysis of amino acid residues essential in the chaperone function of calreticulin. J Biol Chem 2006;281:2338–46.
[45] Eggleton P, Llewellyn DH. Pathophysiological roles of calreticulin in autoimmune disease. Scand J Immunol 1999;49:466–73.
[46] Ling S, Cheng A, Pumpens P, Michalak M, Holoshitz J. Identification of the rheumatoid arthritis shared epitope binding site on calreticulin. PLoS One 2010;5:e11703.
[47] Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis Rheum 1987;30:1205–13.
[48] Ling S, Pi X, Holoshitz J. The rheumatoid arthritis shared epitope triggers innate immune signaling via cell surface calreticulin. J Immunol 2007;179: 6359–67.
[49] Schrag JD, Bergeron JJ, Li Y, Borisova S, Hahn M, Thomas DY, et al. The Structure of calnexin, an ER chaperone involved in quality control of protein folding. Mol Cell 2001;8:633–44.
[50] Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics 2007;23:2947–8.
[51] Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28: 2731–9.

Supplementary data

Table S1

Primers used in this study.

Designation	Nucleotide sequence 5'-3'	Usage
APv	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTV	cDNA synthesis
AUAP	GGCCACGCGTCGACTAGTAC	PCR (full cDNA)
APv2	GACTCAGGACTTCAGGACTTAGTTTTTTTTTTTTTTTTV	cDNA synthesis; 5'RACE (PCR); PCR (full cDNA)
AUAP2	GACTCAGGACTTCAGGACTTAG	5'RACE (PCR)
T7	TAATACGACTCACTATAGGGCGA	sequencing
SP6	CTATTTAGGTGACACTATAGAATAC	sequencing
CnxCrtFW1	ATGTTYGGNCCIGAYAAITG	PCR (partial cDNA)
CnxCrtRV1	CCRTCCATITCYTCRTCCCA	PCR (partial cDNA)
CrtFw1	CTCCAGACCAGCCAGGACGC	PCR (partial cDNA); southern blot
DLCrtFW1	CCTTACCTGCTGCTGCCGGACA	PCR (full cDNA); PCR (partial gene); sequencing
DLCrtFW2	GACACAAGCCGCCGACATGAA	PCR (full cDNA)
DLCrtFW4	GGACATGCTGCCTGCGAAGA	sequencing
DLCrtFW6	CACTCCTGATGCCACCATGT	sequencing
DLCrtFW7	GAGCAGGAGGACATGGAGAG	PCR (partial gene)
DLCrtFW9	GACGACAGAGCCAAGATTGA	PCR (partial gene)
DLCrtFW10	TGAGTACCTTTGCAGAGTCA	PCR (partial gene)
DLCrtFW11	GAAACCTGGGGAGCTACTAAG	PCR (partial gene)
DLCrtFW12	TCACACAGGCAGCGGAGCAA	PCR (partial gene)
DLCrtRV2	GTCGATCCTCACCTCGTATG	sequencing
DLCrtRV4	GTCAGCTCGTCATCCTTGCA	PCR (partial cDNA)
DLCrtRV5	TTCTTCGCAGGCAGCATGTCC	PCR (partial cDNA)
DLCrtRV6	CTGCTCGCGCTTGATGGTGAA	5'RACE (PCR)
DLCrtRV7	CTCTGACCTCGACCAAAGCAACAT	5'RACE (cDNA synthesis); southern blot
DLCrtRV8	AAAACAGTGTGCGCCCACGG	PCR (partial gene); sequencing
DLCrtRV10	TCAGGGTCAGGGATGGTCTC	PCR (partial gene)
DLCrtRV11	GTCGTCGTCCTCCATCTCTT	PCR (partial gene)
DLCrtRV13	TTAGTAGCTCCCCAGGTTTC	PCR (partial gene)
DLCrtRV14	GCTTCCACTCGCCATAGTCT	sequencing
DLCrtRV16	AGTAGAGCCTGACCAATACA	PCR (partial gene)
DLCrtRV18	GCTGGCCATTAGAAAGACTGC	PCR (partial gene)

Letter nucleotide code: H= A, C, T; R= A, G; Y= C, T; S= G, C; W= A, T; V= A, G, C; N= A, T, C, G; I= deoxyinosine.

Table S2

Accession numbers of sequences used in this study.

Species	CRT	CNX
O. niloticus	XP_003448810	-
O. latipes	ENSORLP00000003650	-
O. mykiss	NP_001117950	-
S. salar	ACI32936	-
D. rerio	AAH68336	NP_998613
I. punctatus	AAQ19852	NP_001187109
L. calcarifer	ADQ92842	-
P. olivaceus	ABG00263	-
T. rubripes	ENSTRUP00000024064	-
G. gallus	AAS49610	NP_001025791
X. laevis	NP_001080765	NP_001080326
X. tropicalis	NP_001001253	-
R. rugosa	BAA11425	-
R. norvegicus	NP_071794	-
M. musculus	NP_031617	-
O. cuniculus	NP_001075704	-
B. taurus	NP_776425	NP_001099082
H. sapiens	NP_004334	AAA36125
I. pacificus	AAR29955	-
D. melanogaster	CAA45791	-
A. thaliana	AAC49695	-

Table S3

D. labrax CRT precursor is highly similar to the CRT molecules from different species.

	Similarity (%)	Identity (%)
D.labrax	100.0	100.0
O.niloticus	93.0	84.8
O.latipes	92.0	81.7
O.mykiss	90.4	79.1
S.salar	90.1	79.6
D.rerio	88.3	77.7
I.punctatus	92.3	79.7
L.calcarifer	89.2	77.2
P.olivaceus	90.1	77.8
T.rubripes	89.0	76.1
G.gallus	81.5	67.6
X.laevis	83.8	66.5
X.tropicalis	84.3	68.9
R.rugosa	82.4	67.2
R.norvegicus	83.6	69.0
M.musculus	84.0	69.5
O.cuniculus	85.2	69.7
B.taurus	84.7	69.2
H.sapiens	85.7	71.3
I.pacificus	77.0	63.5
D.melanogaster	74.9	59.5
A.thaliana	71.1	52.2

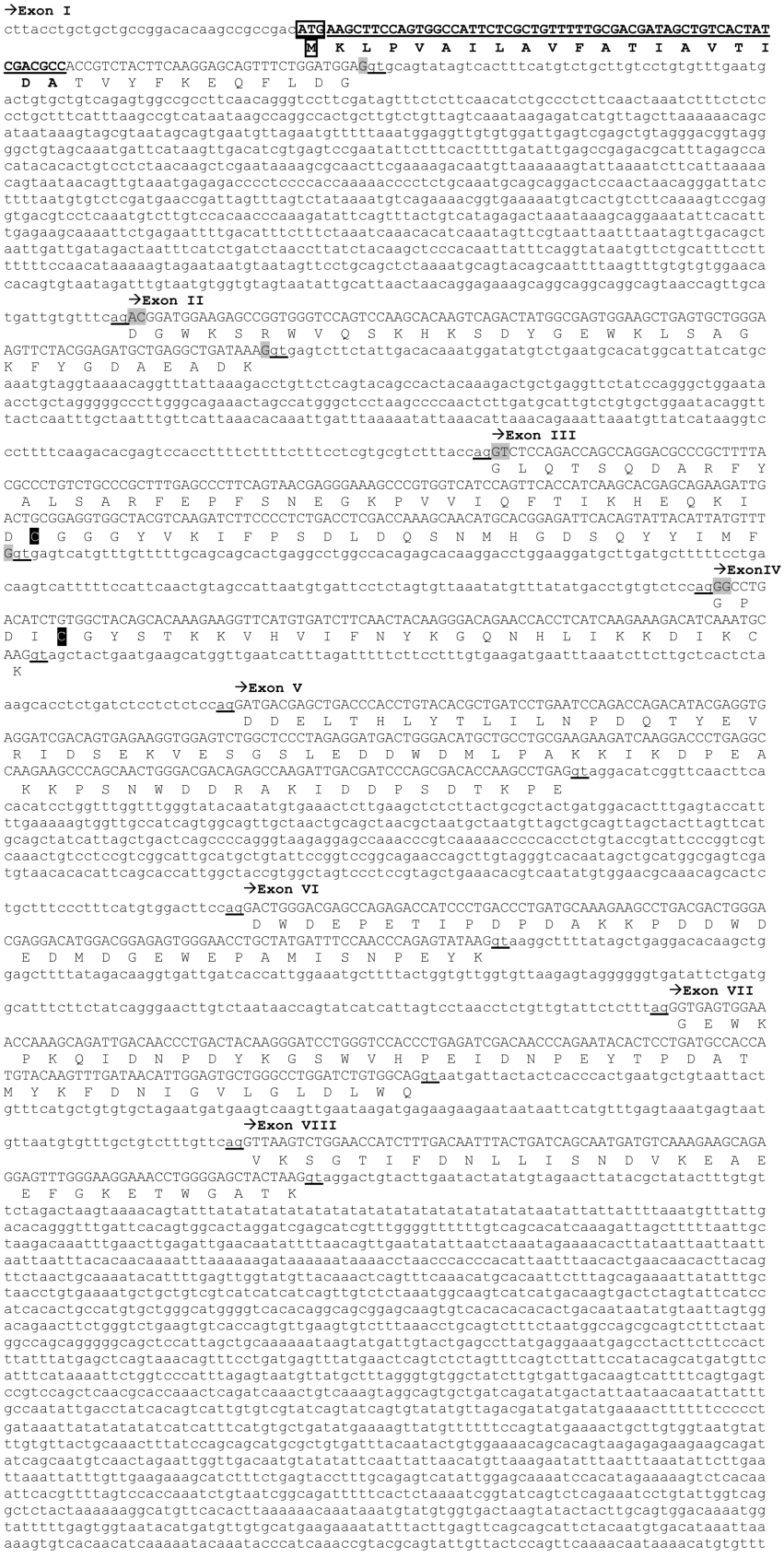


Figure S1. Nucleotide and deduced amino acid sequence of sea bass CRT gene. The ORF and the corresponding predicted protein sequence are both shown in uppercase letters. The translation start and stop codons and the initial methionine are boxed. The predicted signal peptide is highlighted in bold type, and cysteines involved in intramolecular disulfide link are in inverted type. Codons split by phase 1 introns are shaded in grey. The intron splicing consensus sequences (gt/ag) are underlined. Within the 3'UTR, the putative instability motifs are shaded in grey and in bold type.

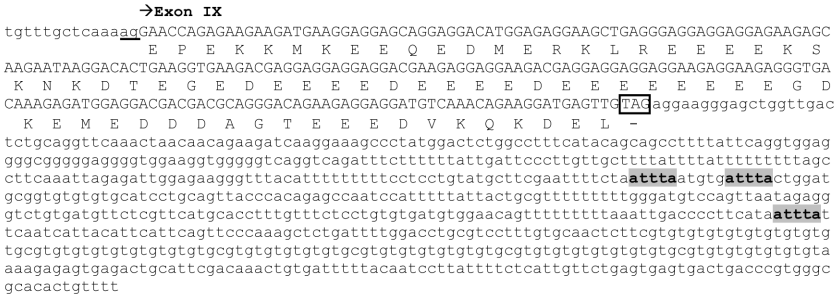


Figure S1 *(continued)*

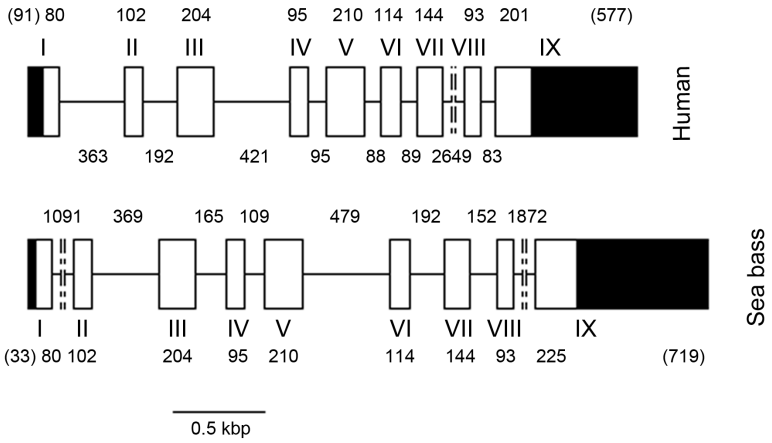


Figure S2. Sea bass and human CRT gene structures. Boxes represent exons and horizontal lines introns. White and black boxes differentiate coding and untranslated regions, respectively. Values above/below boxes and below/above the lines represent the number of nucleotides (size) of exons and introns, respectively. Exon numbers are indicated close to boxes by Roman numerals.

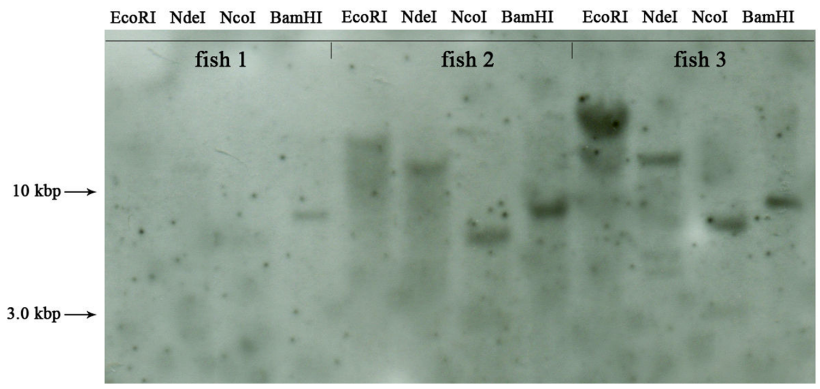


Figure S3. Sea bass CRT is a single copy gene. Southern blot analysis of genomic DNA from three animals digested with four restriction enzymes (EcoRI, NcoI, NdeI or BamHI; all zero cutters within the probe). The blot was later probed with a region of CRT that included part of exon III. A portion of the sea bass calreticulin cDNA (exon III) was amplified with primers CrtFW1/DLCrtRV7 (Table S1), labelled and used as probe. Preparation of the labelled probes, hybridization and post-hybridization stringency washes were performed accordingly to Gene Images™ AlhPhos Direct™ Labelling and Detection System kit (Amersham Biosciences). For signal generation and detection the Chemiluminescent Signal Generation and Detection with CDP-*Star*™ protocol from the same kit was followed. One band can be seen for each digestion. Sizes of the MW marker are indicated on the left.

CHAPTER 6

Two thioredoxin-superfamily members from sea bass (*Dicentrarchus labrax*, L.): characterization of PDI (PDIA1) and ERp57 (PDIA3)

Rute D. Pinto, Ana R. Moreira, Pedro J. B. Pereira and Nuno M. S. dos Santos

Fish & Shellfish Immunol. 2013, *submitted*

Abstract

PDI (PDIA1) and ERp57 (PDIA3), members of the PDI family and of the thioredoxin (Trx) superfamily, are multifunctional proteins with wide physiological roles and have been implicated in several pathologies. Importantly, they are both involved in the MHC class I antigen presentation pathway. This paper reports the isolation and characterization of full cDNA and genomic clones from sea bass (*Dicentrarchus labrax*, L.) PDI (Dila-PDI) and ERp57 (Dila-ERp57). The genes are ~12.4 and ~7.1 kb long, originating 2155 and 2173 bp transcripts and encoding 497 and 484 amino acids mature proteins, for Dila-PDI and -ERp57, respectively. The PDI gene consists of eleven exons and ERp57 of thirteen. As described in other species, both molecules are composed of four Trx-like domains (abb'a) followed by a C-terminal tail, retaining two CGHC active sites and an ER-signalling sequence, suggestive of a conserved function. Additionally, three-dimensional homology models further support Dila-PDI and Dila-ERp57 as orthologs of mammalian PDI and ERp57, respectively. Finally, high similarity is observed to their vertebrate counterparts (> 69% identity), especially among the few ones from closely related teleosts (>79% identity). Hence, these results provide relevant primary data and will enable further studies to clarify the roles of PDI and ERp57 in European sea bass immunity.

1. Introduction

In mammals, the family of protein disulfide isomerases (PDIs) includes 21 related members (reviewed in [1]), abundant in the endoplasmic reticulum (ER) where they play a critical role in protein folding and quality control (reviewed in [2]). However, membership is mostly based in sequence and structural similarity rather than in a common function (reviewed in [3]), therefore including members without redox activity on cysteine residues but that hold relevant chaperone-like activities (reviewed in [4]). The basic unit of PDIs is a thioredoxin-like domain, homologous to the cytosolic reductase thioredoxin (Trx) (reviewed in [5]) that can either be catalytically active (**a**-type) or inactive (**b**-type) (reviewed in [3]). The first typically contain a CXXC active site motif, while the second lack the cysteines making them redox-inactive (reviewed in [3]). Additionally, among PDIs the non-catalytic domains have lower sequence identities than the catalytic ones, showing in some cases structural variations to the Trx fold ($\beta\alpha\beta\alpha\beta\alpha\beta\beta\alpha$) (reviewed in [4]). PDI family members, which have distinct official designations according to the HUGO gene nomenclature committee - HGNC (reviewed in [1], vary in length and domain arrangement and most are soluble proteins with a C-terminal KDEL-like ER-localization signal (reviewed in [3, 6]).

PDI itself (PDIA1 according to HGNC) is the founding member of the PDI family (reviewed in [3]), the first known to act as a catalyst of disulfide bond formation [7]. Besides its critical role in the formation, breakage and rearrangement of disulfides (reviewed in [8, 9]) and as a chaperone [10], it is also the subunit of distinct protein complexes (reviewed in [4]), being actually recognized as a versatile and multifunctional protein that participates in a variety of processes at distinct cellular locations (reviewed in [11]). Not surprisingly, PDI is implicated in several pathophysiologic conditions such as neurodegenerative diseases, autoimmunity and cancer (reviewed in [11]). PDI has a [abb'xa'c] domain organization, where **a**, **b**, **b'** and **a'** are the four Trx-like domains, **x** is a 19-residue linker connecting **b'** to **a'**, and **c** is an acidic carboxyl tail [12-14]. A CGHC redox-active site is present at each of the **a** and **a'** domains, where the upstream cysteine forms mixed disulfides with substrate proteins, while the downstream one is involved in the escape pathway [15]. The **b'** domain harbours a hydrophobic pocket, which is the major substrate binding site [16, 17]. Two yeast PDI crystal structures (determined at 4 and 22 ºC) unveiled a highly flexible protein, where the catalytic domains represent two mobile arms connected to a non-catalytic rigid base, allowing binding of substrates of various sizes and conformations [13, 18]. In the twisted 'U form', the active sites face each other at opposite ends of the U with the hydrophobic patches on **b** and **b'** domains buried

inside the U [13], while in the 'boat form' the molecule adopts a more open conformation in which the catalytic domains do not face each other [18]. Despite the considerable amount of information available, little is known regarding PDI's activities *in vivo*, reflect of its likely essential nature and broad substrate specificities [3].

First isolated in 1988 [19], ERp57 (according to HGNC) is a homolog of PDI (reviewed in [20]) and one of the family members with highest identity to PDI [1]. Mainly existing in the ER, but also localized in the nucleus [21], cytoplasm [22] and cell surface [23-25], ERp57 has been linked to several physiological (e.g. modulation of platelet function [26]) and disease states, such as cancer [23] and Alzheimer's (reviewed in [20]). Similarities to PDI include sequence identity (~50% in **a** or **a'** domains; ~20% in **b** or **b'** domains), domain organization, presence of a CGHC motif at the active sites (reviewed in [20]), and a strikingly similar overall U shape, as shown by both solution and crystal structures [27, 28]. On the other hand, distinctive features comprise a highly charged region within the **b'** domain (no hydrophobic pocket) [27], which is the calnexin binding site, localized on the opposite side of the PDI substrate binding site (reviewed in [29]), and a shorter basic C-terminal domain (reviewed in [20]). As PDI, it is a multifunctional oxireductase, although catalyzing at a lower rate thiol oxidation, reduction and isomerisation *in vitro* [30] and possibly in ER as well [31]. Contrary to PDI, ERp57 specifically acts on glycosylated substrates through interaction with the lectin-like chaperones calnexin (CNX) and calreticulin (CRT) [32]. Furthermore, ERp57 is most widely known for its involvement in the immune system, due to the critical role in MHC class I assembly (reviewed in [33]). At an early stage, ERp57 and CNX assist disulfide bond formation of glycosylated MHC class I heavy chain (HC), enabling subsequent dimerization with soluble β2m [33]. This early oxidative folding of HC is delayed upon ERp57 inactivation [34]. At a late stage, the empty heterodimer enters the peptide-loading complex (PLC; CRT/ERp57/TAP/TPN), in order to acquire peptide ligands [33]. ERp57 binds TPN through its catalytic domains [28] and despite formation of a covalent bond [35], the interaction seems to be stabilized by cumulative protein-protein contacts (reviewed in [4]). Additionally, ERp57 may function independently of its structural role in the PLC, specifically reducing α2 in released sub-optimally loaded/unstable complexes (reviewed in [3]). As ERp57 [35, 36], PDI has been implicated in class I assembly as well [37]: it facilitates optimal peptide binding, maintaining HC oxidized state (re-oxidation of α2) [37]; and serves as a redox-regulated peptide carrier, delivering peptides from TAP to HC [38]. PDI has been further involved in the regulation of PLC disassembly [39].

Previously we have reported the characterization of European sea bass molecules [40-43] described in the literature to be involved in MHC class I antigen presentation and

processing. Here we report the isolation of sea bass PDI and ERp57 full genes and transcripts with further *in silico* analysis at both structural and phylogenetic levels. Collectively, these studies will enable the development of functional studies of class I pathway in this relevant fish species.

2. Materials and methods

2.1. Fish

Sea bass, *Dicentrarchus labrax*, were kept in a recirculating, ozone-treated salt-water (20-25‰) system at 22 ± 1 ºC and fed with commercial pellets twice a day. Fish were sacrificed with a lethal dose of 2-phenoxyethanol (Panreac; >0.5 mL/L).

2.2. cDNA cloning

RNAs were transcribed to cDNAs following the BioScript RNase H^- (Bioline) protocol. PCR products were purified and cloned into pGEM-T Easy (Promega). Automated sequencing was performed using primers detailed in Supplementary Table S1. 5'RACE experiments were performed following the 5´RACE System v. 2.0 (Invitrogen) and Recombinant Terminal Transferase (Fermentas or Roche) was used to dATP-tail the purified cDNA.

2.2.1. PDI

A partial genomic sequence of sea bass PDI retrieved from the EMBL gene database ([http://www.ebi.ac.uk/genomes/]: [CABK01014256]; contig ctg479172) was used to design three specific primers for direct use in 5'RACE. To obtain the 5' untranslated region (UTR), a 5'RACE was performed using cDNA synthesised with DLPDIRV2 (Table S1) using total RNA extracted from the liver of a non-stimulated fish, as described above. The tailed cDNA was first amplified with primers APv/DLPDIRV3 (Table S1) followed by a second PCR using primers AUAP/DLPDIRV4 (Table S1). The ~850 bp fragment obtained was purified, cloned and sequenced.

To obtain the full sea bass PDI transcript, specific primers were designed at the beginning of the 5' UTR region. The same initial liver total RNA was reverse transcribed using primer APv (Table S1). Primers DLPDIFW2/AUAP (Table S1) were used in a first amplification PCR followed by a second PCR step using the same primers or the set DLDIFW3/AUAP (Table S1). The ~2200 bp fragment obtained was purified, cloned and 4 independent clones sequenced.

2.2.2. ERp57

Total RNA from the spleen of one fish was reverse transcribed with primer APv (Table S1). Degenerate primers were designed on conserved regions of ERp57 across several vertebrate species. The spleen cDNA was amplified with primers ERp57FW1 and ERp57RV1 (Table S1). A semi-nested PCR was performed with primers ERp57FW1/ERp57RV2 and the obtained product of ~700 bp was purified, cloned and sequenced.

To reach the 5' UTR, two consecutive 5'RACE experiments were performed: (i) total RNA extracted from the spleen of one fish was reverse transcribed with primer DLERp57RV5 and the tailed cDNA was first amplified with APv/DLERp57RV4 followed by a PCR with AUAP/DLERp57RV3 (Table S1); (ii) total RNA extracted from the head kidney of one fish was reverse transcribed with primer DLERp57RV7 and the tailed cDNA was first amplified with APv2/DLERp57RV7 (Table S1), followed by an amplification with AUAP2/DLERp57RV6 (Table S1). PCR products of ~400 (i), and ~300 (ii) were consecutively obtained, purified, cloned and sequenced.

To obtain the full sea bass ERp57 cDNA, specific primers (Table S1) were designed at the beginning of the 5' UTR region. DLERp57FW9 and AUAP were used in an amplification using the same non-stimulated liver cDNA used for PDI isolation. The obtained ~2200 bp PCR product was purified, cloned and 3 independent clones were sequenced.

2.3. Genomic DNA cloning

Genomic DNA (gDNA) was isolated from sea bass erythrocytes from a single fish, as previously described [44]. The full PDI gene was obtained through five separate PCR amplifications using the following sets of primers: DLPDIFW2/DLPDIRV9, DLPDIFW6/DLPDIRV2, DLPDIFW10/DLPDIRV10, DLPDIFW7/DLPDIRV8 and DLPDIFW5/DLPDIRV6 (Table S1). The PCR products of ~4.5, ~2.5, ~0.6, ~1.4 and ~4.0 kb, respectively, were purified, cloned and one clone of each product sequenced as before. The full PDI gene was assembled from the overlapping partial sequences.

To obtain the full ERp57 gene, gDNA was amplified by PCR with DLERp57FW8/DLERp57RV11 (Table S1) followed by a nested PCR with primers DLERp57FW9/DLERp57RV12 (Table S1). The PCR product (~7.0 kbp) was purified, cloned and one clone sequenced as before.

2.4. Sequence analysis

Full nucleotide and protein sequences from sea bass PDI and ERp57 molecules were compared to several PDI and ERp57 sequences currently available in the GenBank database [http://www.ncbi.nlm.nih.gov/genbank/]. The multiple sequence alignment was made using CLUSTALW [45] and formatted with Bioedit [46]. Two Trx domain profiles as well as their respective Trx-family active sites have been identified with Prosite [47]. The Trx domain boundaries were based on human PDI [14]. The potential nuclear localization signal was predicted at PredictProtein [www.predictprotein.org]. The molecular weights of the polypeptide chains were calculated with the Expasy compute pI/MW tool [http://www.expasy.org]. Putative N-glycosylation sites and signal peptide sequences were predicted using NetNGlyc 1.0 [http://www.cbs.dtu.dk/services/NetNGlyc] and SignalP 4.1 [http://www.cbs.dtu.dk/services/SignalP] servers, respectively. Hydrophobicity was predicted with Protscale [http://web.expasy.org/protscale/] and drawhca [http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=HCA]. The Neighbour-joining phylogenetic tree was constructed with MEGA version 5.05 [48], using p-distance parameter and pairwise deletion of gaps, and tested for reliability using 1000 bootstrap replications. The percentages of similarity and identity were calculated with MatGAT [49] using default parameters.

2.5. Southern blotting

Genomic DNA was digested with different restriction enzymes (EcoRI, NcoI, BamHI and HindIII for ERp57; all zero cutters within the probe), separated in a 0.8% agarose gel and subjected to Southern blotting [50]. A portion of sea bass ERp57 cDNA (exon I) was amplified with primers DLERp57FW8/DLERp57RV6 (Table S1), labelled, and used as probe. Preparation of the labelled probe, hybridization and post-hybridization stringency washes were performed accordingly to Gene Images AlhPhos Direct Labelling and Detection System (Amersham Biosciences) kit. For signal generation and detection the Chemiluminescent Signal Generation and Detection with CDP-*Star* protocol from the same kit was followed.

2.6. 3D modelling

Suitable structural templates for sea bass PDI (human PDI from PDB entries 1mek [51] and 3uem [14]) were identified by a BLAST search as implemented in the SWISS-MODEL Protein Modelling Server [52]. Homology modelling was performed with SWISS-MODEL in automated mode [52] with 80% and 74.5% identity between target and template molecules, respectively. The only gap introduced (3uem alignment) involves secondary structure elements (SSE). The resulting theoretical models of sea bass PDI **a**

domain alone, and of **bb'a'** domains, have calculated global energies of -1739.826 kJ/mol and -10853.179 kJ/mol, respectively. By superposing both homology models with human ERp57 (PDB entry 3f8u [28]), a full-length model of Dila-PDI was generated. SSE for the models were computed with PROMOTIF [53]. Model quality was assessed with WHATCHECK [54] and PROCHECK [55] as implemented in the Swiss Model structure assessment tool. Model representations were prepared with PyMOL (http://pymol.org/).

Similarly, a suitable structural template for sea bass ERp57 (human ERp57 from PDB entry 3f8u [28]) was identified by a BLAST search as implemented in the SWISS-MODEL Protein Modelling Server [52]. Homology modelling was also performed with SWISS-MODEL in automated mode [52]. The calculated global energy of the model is -19239.387 kJ/mol. The alignment shows high identity (~77%) between the target and structural template molecules with one gap introduced outside the SSE (data not shown). The SSE were determined with PROMOTIF [53]. Model quality structural assessment and preparation of model figures were performed as previously described for PDI.

3. Results and discussion

3.1. cDNA, gene and Southern blot analysis

3.1.1. PDI

The sea bass PDI (*Dila-PDI*) full-length cDNA has been obtained through homology cloning, by performing RT-PCR on mRNA extracted from the liver of one fish. BLAST analysis confirmed highest similarity with fish PDI. The here-reported *Dila-PDI* sequence [GenBank: JX891476] has a total size of 2155 bp, containing a 5' UTR of 161 bp, an ORF of 1542 bp and a 3' UTR of 452 bp (excluding the poly(A) tail). Within the 3' UTR no mRNA instability motifs were detected [56] but two potential non-canonical polyadenylation signals could be identified (Fig. S1A).

In the *Dila-PDI* gene [Genbank: JX891477], which is 12386 bp long, intron donor/acceptor sites follow the gt/ag consensus rule (except for intron 4 - gt/gg), and codons in the exon boundaries are split by phase-one (1, 2 and 8) or phase-zero (3, 4-7, 9 and 10) introns (Fig. 1A). The different regions encoded by each exon are described in Table S2. Some regions are encoded by more than one exon, e.g. each domain is encoded by three different exons, while the x linker and the C-terminal portion are encoded by two exons. The sea bass PDI eleven-exon genomic structure (Fig. 2A) is identical to that of the human gene [57]. Intron sizes are highly variable, but intron phase and location are conserved between sea bass and human PDI molecules, the same being

true for exon sizes (exception for exons I, X, and XI) (Fig. 2A).

A

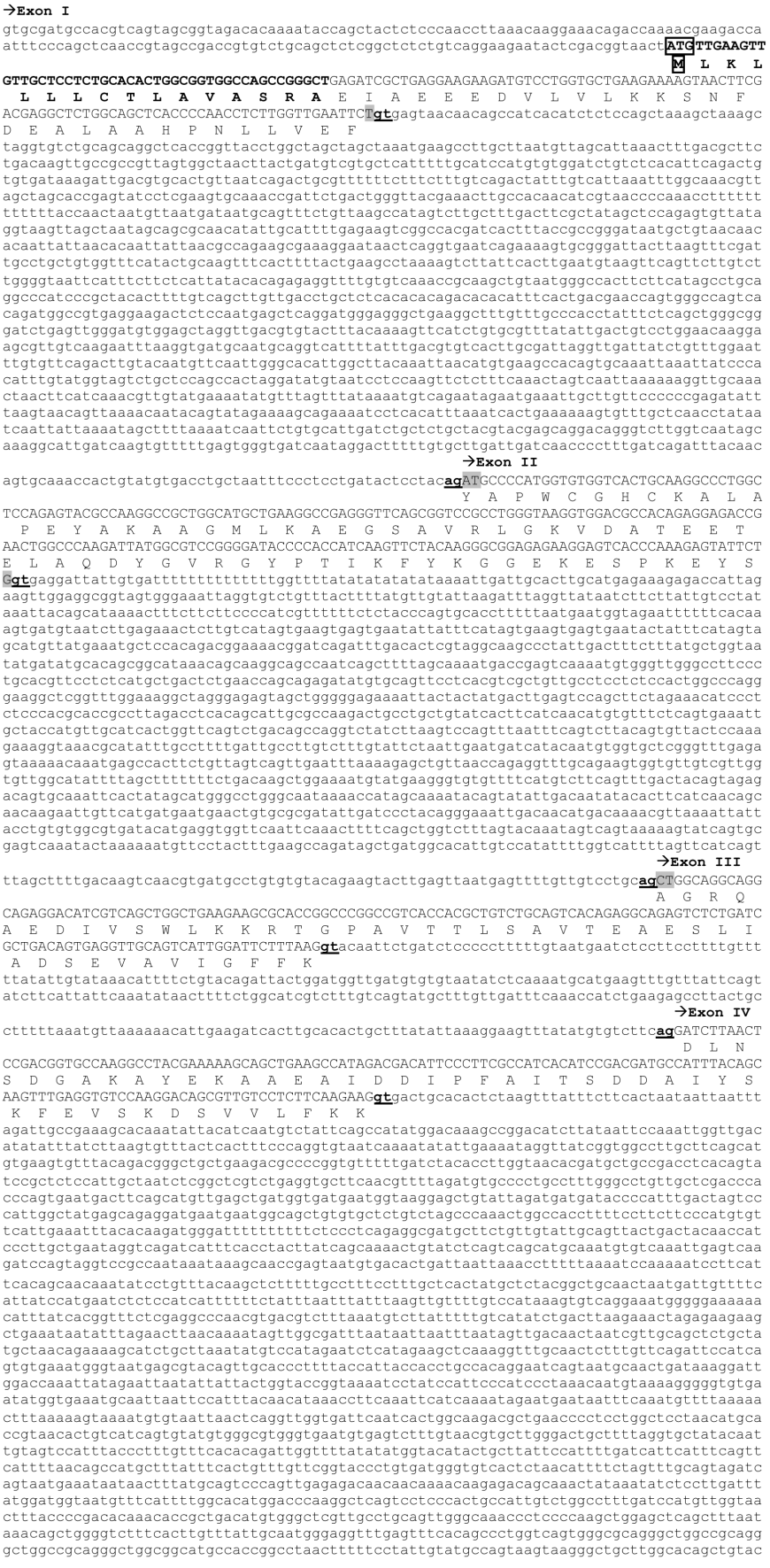


Figure 1. **Nucleotide and deduced amino acid sequences of sea bass PDI and ERp57 genes.** The ORF and the corresponding predicted protein sequences of sea bass PDI (A) and ERp57 (B) are shown in uppercase letters. The translation start and stop codons and the initial methionine are boxed. The predicted signal peptide is highlighted in bold type. Codons split by phase 1 or 2 introns are shaded in grey. The intron splicing consensus sequences (gt/ag) are underlined. Within the 3'UTR, the putative instability motifs are underlined while poly-adenylation signal sequences are shaded in grey and in bold type. Exons are identified above the sequences.

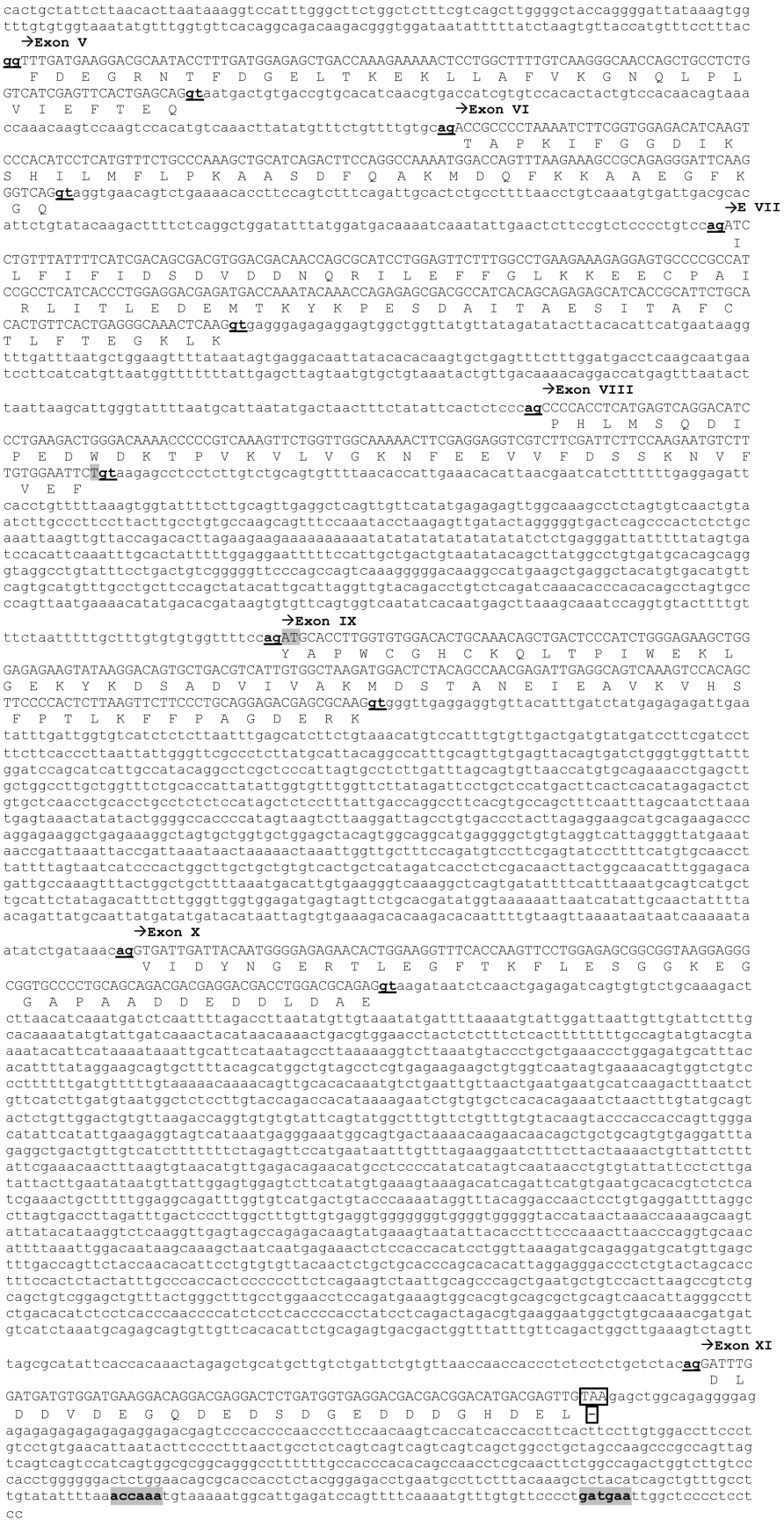


Figure 1 *(continued)*

B

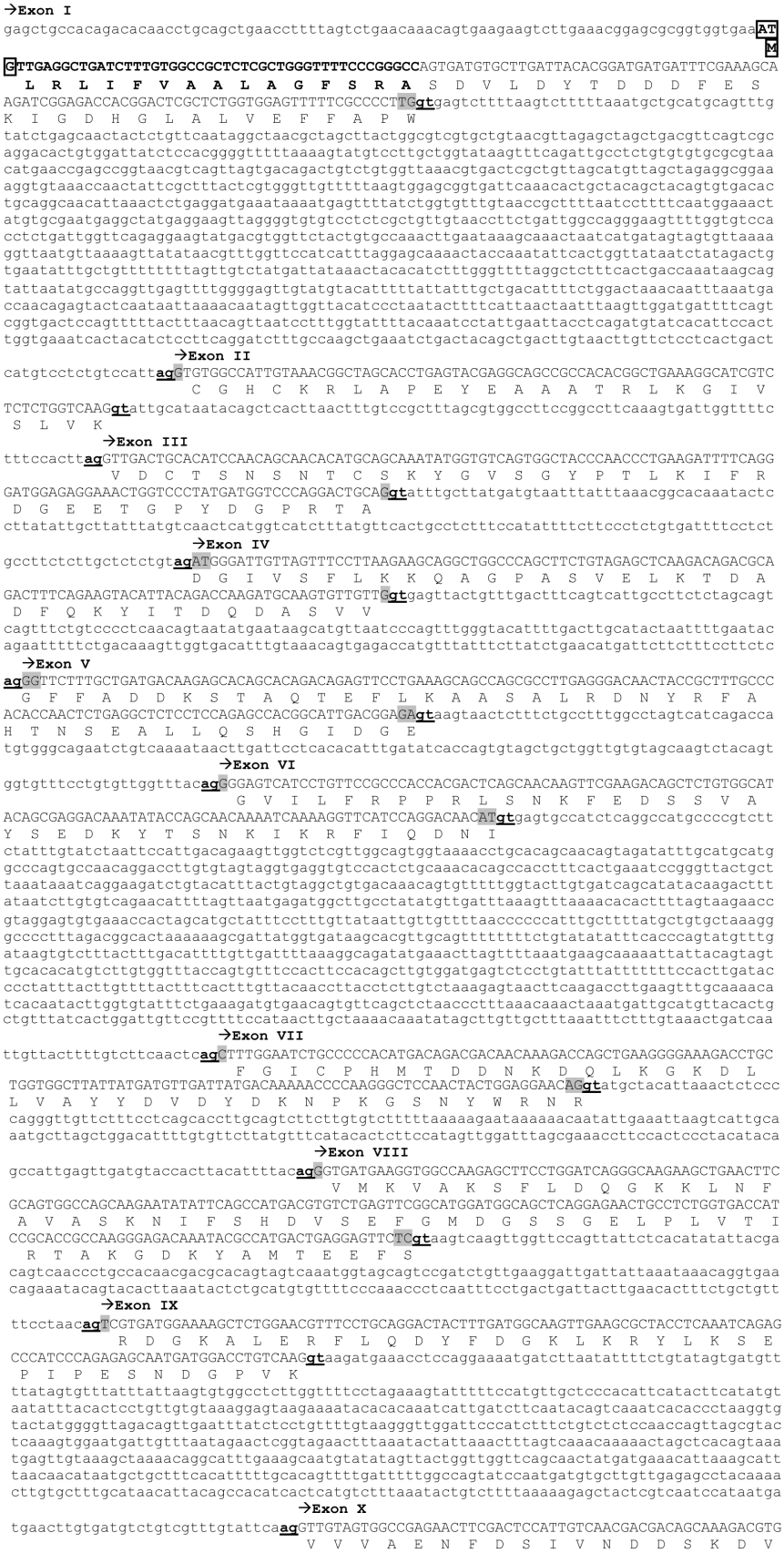


Figure 1 *(continued)*

Figure 1 *(continued)*

3.1.2. ERp57

Similar to *Dila-PDI*, the sea bass ERp57 (*Dila-ERp57*) full-length cDNA was obtained by homology cloning. BLAST analysis confirmed highest similarity with the few fish ERp57 molecules described. The *Dila-ERp57* sequence [GenBank: JX891474] has a total length of 2173 bp, containing a 5' UTR of 164 bp, an ORF of 1503 bp, and a 3' UTR of 506 bp (excluding the poly(A) tail). Within the 3'UTR, one mRNA instability motif [56] and three potential polyadenylation signals were identified (Fig. S1B).

In *Dila-ERp57* gene [Genbank: JX891475] that is 7119 bp long, all intron donor/acceptor sites follow the gt/ag consensus rule, and codons in the exon boundaries are split by phase-two (1, 5-8 and 11), phase-one (3, and 4) or phase-zero (2, 9, 10 and 12) introns (Fig. 1B). The typical domains of the protein are encoded by three to five exons (Table S2), while the x linker, C-terminal tail, UTRs, and signal peptide coding information are present in single exons. The sea bass ERp57 has a thirteen-exon genomic structure, identical to those of human [58] (Fig. 2B) and mouse [24] (not shown). Conservation of exon sizes (except exons I and XIII), intron phase and location is observed between sea bass and human ERp57 (Fig. 2B).

Southern blot analysis revealed that *Dila-ERp57* is a single-copy gene (not shown), which is in agreement to what has been described for human [58] and mouse molecules [24]. Among salmonids, however, two copies of ERp57 appear to be present [59, 60].

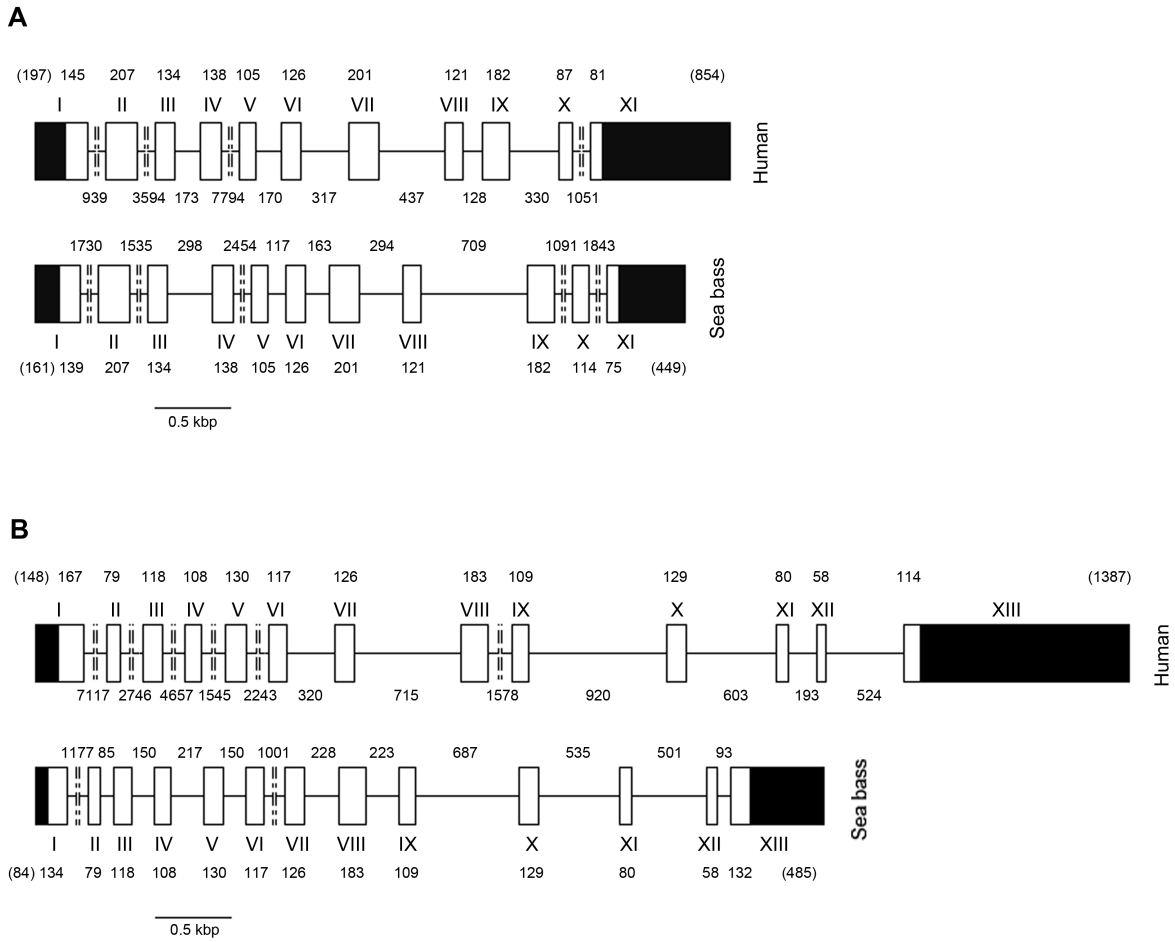


Figure 2. Comparison between sea bass and human PDI and ERp57 gene structures. Boxes represent exons and horizontal lines introns of PDI (A) and ERp57 (B). White and black boxes differentiate coding and untranslated regions, respectively. Values above/below boxes and below/above the lines represent the number of nucleotides (size) of exons and introns, respectively. Exon numbers are indicated close to boxes by Roman numerals.

3.2. Primary structure analysis

3.2.1. PDI

Amino acid sequences of PDI (PDIA1) from several species (Table S3) were aligned with that of Dila-PDI (Fig. 3). In sea bass, a predicted 497-residue (55.1 kDa) mature protein would be generated by cleavage of the putative 16-amino acid signal peptide. Variation in polypeptide size occurs across species, Dila-PDI being the second largest protein of all analysed species, essentially due to a long acidic (E/D-rich) C-terminal region (Fig. 3).

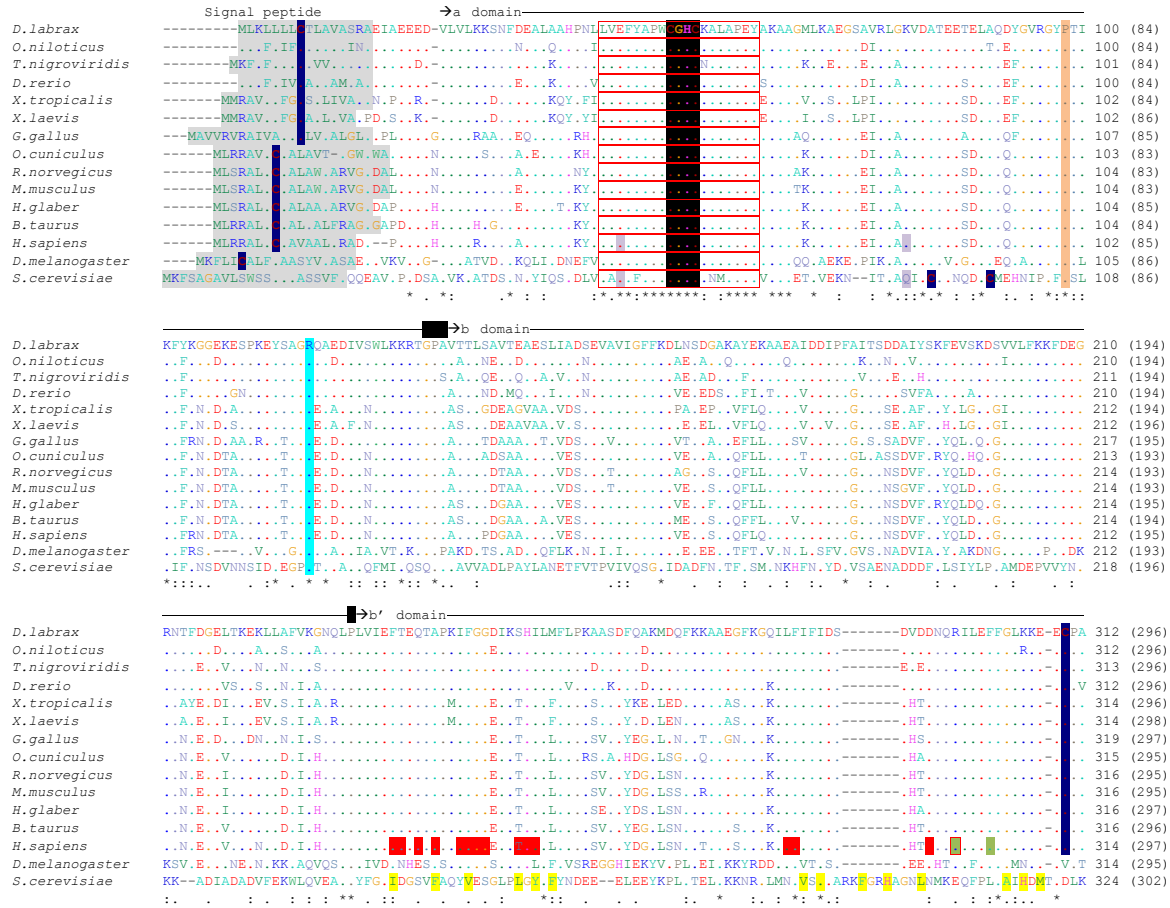


Figure 3. Alignment of PDI amino acid sequences. Amino acid sequences were retrieved from GenBank (Table S3) and aligned with CLUSTALW [45]. Dashes indicate gaps that maximize the alignment and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with (*), (:) and (.), respectively. Residue numbers are given to the right of the sequences, with numbers in parenthesis referring to the mature proteins. Signal peptides are shaded in grey and denoted above the alignment. Predicted **a**, **b**, **b'**, and **a** domains based on human PDI [14], are indicated above the alignment, with junction regions represented by black boxes. The two Trx signature motifs of domains **a** and **a'** are boxed, with each active site shaded in black. Non-catalytic cysteines are shaded in dark blue. Residues possibly involved in triggering PDI conformational changes, including cis-prolines near each active site [13], are shaded in orange [14]. Arginine residues relevant for reoxidation are shaded cyan [70]. Two buried polar residues related to the redox state [13] are shaded in purple. Hydrophobic patch on the b' domain putatively involved in peptide/substrate binding is shaded yellow [13]. Contacts of x-linker with **b'** or **a'** are shaded in light salmon; direct contacts between **b'** and **a'** domains are shaded in green [14]. Residues identified as part of the substrate binding site of **b'** domain are shaded in red [16, 66]. One low complexity region at the C-terminal extension is underlined and in bold type. KDEL-like ER-retention signals are boxed.

Sea bass PDI contains the four typical Trx domains followed by a C-terminal tail, overall organized as abb'xa'c (Fig. 3). Within **a** and **a'** domains, two Trx superfamily signature motifs (L^{43}VEFYAPWCGHCKALAPEY and F^{387}VEFYAPWCGHCKQLTPIW) were identified in Dila-PDI, both consisting of highly conserved regions (Fig. 3). Each of

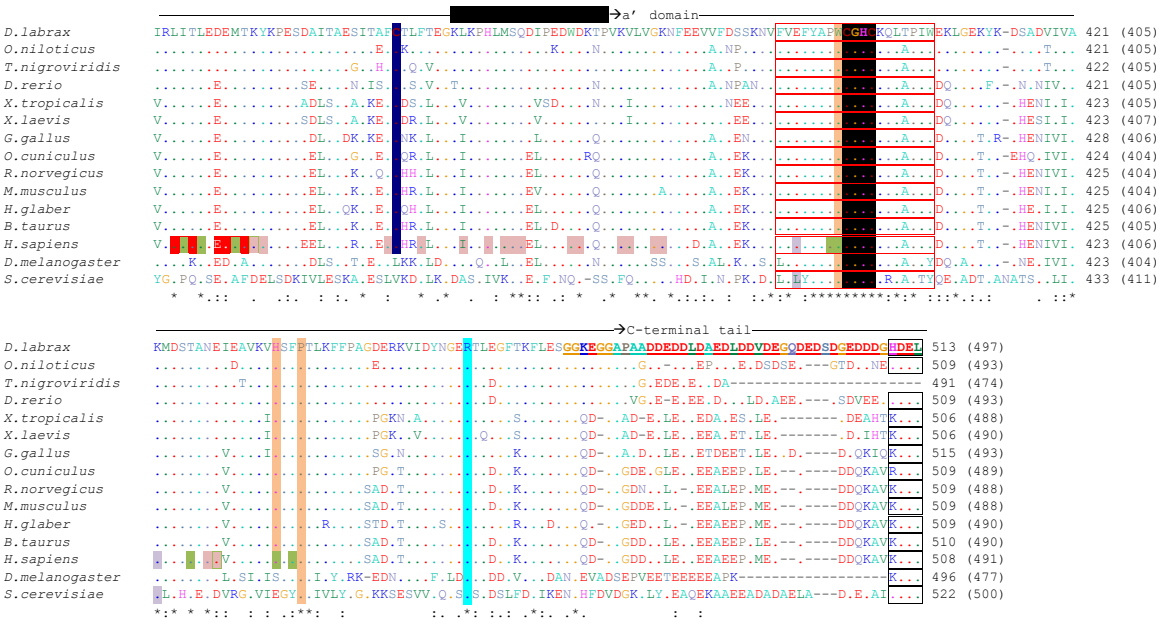


Figure 3 *(continued)*

these motifs contains two conserved cysteine residues (C^{51} and C^{54}; C^{395} and C^{398}), which are part of the CGHC active site (Fig. 3). This site is important for efficient oxireductase activity, and any individual enzyme that contains it along with a **b'**-like domain is capable of catalyzing oxidation, reduction and isomerisation of disulfide bonds (reviewed in [29]). Two non-active site cysteines near the N-terminus of yeast PDI (absent in all vertebrate species) form a disulfide [61] that is highly stable and affects **a** domain redox activity [62], but the two additional cysteines in the **b'** domain of mammalian PDI (conserved in all analysed vertebrates) are present as thiol groups, and have no influence in enzymatic activity [63]. The redox potentials of **a** and **a'** domains in mammalian PDI are similar [64], a property thought to be related with conservation of two residues near each active site: E^{47}/K^{81}, E^{391}/K^{422} in *H. sapiens* [13]. In yeast PDI (yPDI), E^{55}/Q^{87} and L^{400}/K^{434} are present instead, hence, **a** and **a'** domains have asymmetric redox states [62] (and accessibilities), both consistent with the experimental three-dimensional structure [13]. Since all other species show conservation of the residues present in hPDI (hPDI; Fig. 3), symmetric redox potential of the catalytic domains is expected. According to a recent study on hPDI, cysteines in the active sites are involved in a redox-regulated mechanism to control hPDI chaperone activity (higher when oxidized) [14]. Apparently, oxidized hPDI adopts a loose conformation, whereas the reduced molecule is very compact, suggesting that different redox conditions induce significant conformational changes that affect chaperone activity [14]. Conservation of the CGHC motif at both Dila-PDI active sites, suggests oxireductase as well as redox-regulated chaperone activities might be evolutionarily conserved. The **b'** domain is the core/primary site for substrate binding, but contiguous portions of the other domains also contribute to this binding [65]. In yPDI, a continuous hydrophobic surface

formed by exposed hydrophobic patches is suggested to be crucial for the interaction with its substrates [13]. Despite the lack of perfect sequence conservation across species (Fig. 3), conserved patches of hydrophobic residues are similarly observed (not shown). In the reduced form, the **a'** domain extensively packs with both the **b'** domain and the x-linker, shielding substrate-binding areas, which explains the lower chaperone activity of reduced hPDI [14]. Contacts between the x-linker and the **b'** or **a'** domains occur with a similar pattern in yPDI, ERp57 and reduced bb'xa' (hPDI: K^{326}, Y^{327}, F^{342}, F^{346} at **b'** domain; I^{351}, H^{354}, M^{356}, S^{357}, Q^{358}, W^{364}, D^{365} at x-linker; K^{370}, V^{371}, G^{374}, K^{375}, N^{430}, E^{431} at **a'** domain), but direct interactions between domains **b'** and **a'** are mainly observed in reduced bb'xa' (hPDI: R^{300}, F^{304}, I^{318}, L^{320}, M^{324}, K^{326} at **b'** domain; P^{395}, W^{396}, T^{428}, E^{431}, H^{438}, F^{440} at **a'** domain) [14]. Considerably reduced activity with substrates was observed for mutations of the conserved P^{441} (located near the **a'** domain active site in the 3D structure) [13], implicating it, together with H^{438} and W^{396}, in sensing the redox switch of the a´ domain active site and triggering the conformational changes [14]. All these residues are conserved in Dila-PDI as well as in the considered vertebrates (Fig. 3); some were previously identified for substrate binding, namely R^{300}, I^{318}, M^{324} [16] and Y^{327} [66], but further residues were also indicated for substrate binding: T^{241}, Q^{243}, A^{245}, I^{248}, F^{249}, G^{250}, G^{251}, T^{255}, H^{256}, I^{257}, L^{287}, F^{288}, D^{297}, L^{317}, I^{318}, T^{319}, E^{321}, E^{322}, T^{325} [16, 66]. Only T^{255} is substituted for a serine in teleosts, all other residues being conserved across vertebrates (Fig. 3). As other PDI molecules (reviewed in [1]), the C-terminal extension of Dila-PDI has an E/D-rich Ca^{2+}-binding region (Fig. 3); furthermore, these negatively charged residues were shown to play an important role in the activity of yPDI [13].

3.2.2. ERp57

Precursor sequences of ERp57 (PDIA3) molecules from different species (Table S3) were also aligned with that of sea bass (Fig. 4). The Dila-ERp57 displays a 484-residue mature form (54.2 kDa) after cleavage of a 16-residue potential leader peptide. It is the longest molecule among all analysed species (Fig. 4). Similar to other ERp57 molecules, Dila-ERp57 comprises the same basic domain organization as PDI, i.e., two catalytic domains separated by two non-catalytic ones, x-linker and C-terminal region (abb'xa'c) (Fig. 4). At the sequence level, both sea bass molecules show higher identity between **a** domains (~45% for both) and lower among **b** domains (~25 and ~11% for **b** and **b'**, respectively), as described for mammals (reviewed in [20]). Throughout the molecules some indels are also observed (Fig. S2). Each catalytic domain bears a Trx signature motif (L^{38}VEFFAPWCGHCKRLAPEY and L^{387}IEFYAPWCGHCKNLEPKF), with a CGHC active site, conserved across species (Fig. 4). However, slightly distinct redox potentials

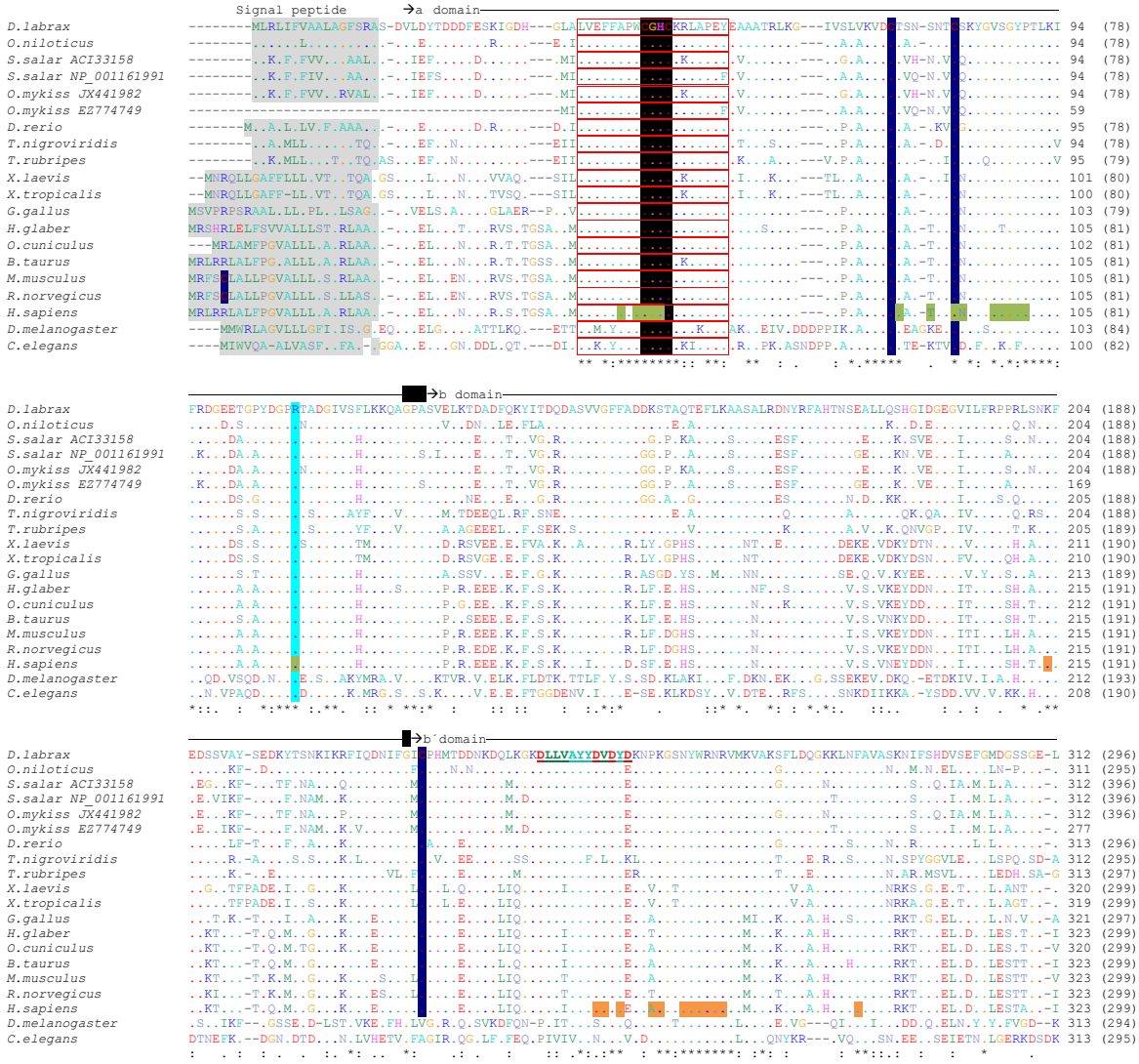


Figure 4. Alignment of ERp57 amino acid sequences. Amino acid sequences were retrieved from GenBank (Table S3) and aligned with CLUSTALW [45]. Dashes indicate gaps that maximize the alignment and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with (*), (:) and (.), respectively. Residue numbers are given to the right of the sequences, with numbers in parenthesis referring to the mature proteins. Signal peptides are shaded in grey and denoted above the alignment. Predicted **a**, **b**, **b'**, and **a** domains, based on human PDI [14], are indicated above the alignment, with junction regions represented by black boxes. The two Trx signature motifs of domains **a** and **a'** are boxed, with each active site shaded in black. Non-catalytic cysteines are shaded in dark blue. Arginine residues relevant for reoxidation are shaded cyan [70]. Residues of human ERp57 that contact CNX/CRT are shaded orange. Residues involved in TPN binding are shaded green (**a** domain) and dark salmon (**a´** domain). Two low complexity regions, at **b'** domain and the C terminus extension, are underlined and in bold type. Putative nuclear localization signal is shaded in sepia. KDEL-like ER-retention signals are boxed.

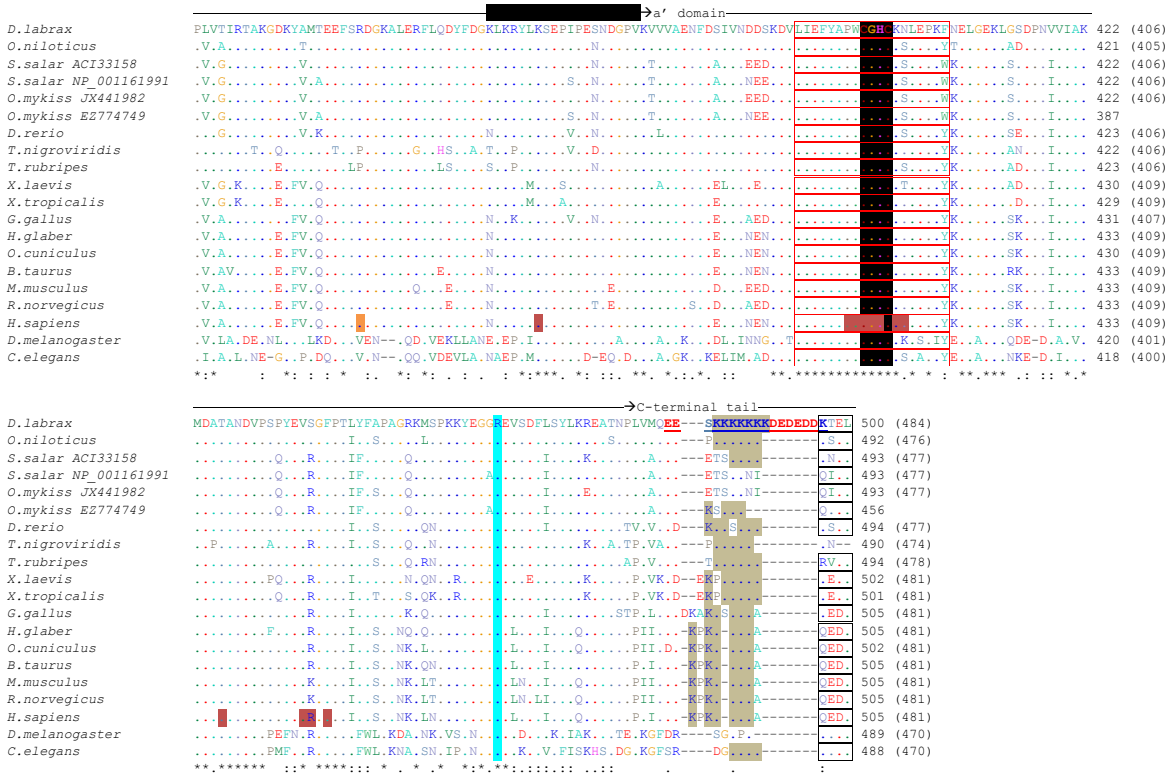


Figure 4 *(continued)*

have been described for hERp57 **a** and **a'** active sites, with values comparable to that of PDI [30]. Two additional cysteines (sbC^{74} and C^{81}) forming a structural disulfide within the **a** domain [30] are observed (similarly to yPDI, Fig. 3, 4 and S3). Furthermore, a third residue (sbC^{233}) is present in the ERp57 **b'** domain of all vertebrates (Fig. 4). The **a'** domain cysteine residues (and their redox state) were reported to play a critical role in the control of the DNA-binding activity of ERp57 [67].

ERp57, but not PDI, has been shown to bind CNX and CRT [68], specifically through **b** and **b'** defined regions [27, 69]. The human ERp57 residues suggested by mutagenesis studies to be involved in CRT binding (including V^{267}, Y^{269}, F^{299}, K^{274}, R^{280} and N^{281}) [69], correlate well with those identified by NMR and mutagenesis studies as part of the CNX binding site (K^{214}, D^{266}, V^{267}, A^{273}, K^{274}, N^{277}, Y^{278}, W^{279}, R^{280}, N^{281}, R^{282} and R^{344}) [27, 69]. Among these 14 residues implicated in human ERp57 interaction with CNX/CRT, 12 are identical while 2 are conservatively substituted in Dila-ERp57 (Fig. 4), suggesting similar interactions may occur. Of note, the main contacts occur through the **b'** domain, whereas the **b** domain provides additional contacts strengthening the interaction [27]. An arginine, located at the β5-α4 loop of active domains, critical for the catalytic function, especially reoxidation, in all human PDIs [70], is present in both Dila-PDI and Dila-ERp57 as well as across species (Fig. 3, 4, S2 and S3).

ERp57's crucial involvement in the folding process of MHC class I has been indicated by detection of disulfide-bonded intermediates with class I HC [71, 72]. The N-terminal cysteines from both active sites have been associated with ERp57 reductase activity on partially folded (non-β2m associated) HCs *in vitro* [72]. ERp57 activity independent from CNX/CRT has been reported in the class I pathway both as a catalyst and as a member of the PLC [73]. However, recent evidence has been provided for CRT-mediated recruitment of MHC through the detection of HC/β2m/CRT intermediate before PLC engagement, which was strictly glycan-dependent [74]. ERp57 associates covalently with TPN [75], through formation of a disulfide bond between TPN C^{95} and the N-terminal cysteine at the redox-active site of the **a** domain (hERp57 C^{57}), which is conserved across species (Fig. 4). Furthermore, this stable association is reinforced by multiple protein-protein interactions [35]. Accordingly, mutation of the cysteines of the **a'** domain active site (not covalently linked to TPN) significantly diminished the efficacy of complex formation [75]. Contacts at TPN interface [28] include 15 residues from the **a** domain (hERp57 A^{54}, W^{56}, C^{57}, G^{58}, H^{59}, T^{86}, T^{89}, C^{92}, N^{93}, V^{97}, S^{98}, G^{99}, Y^{100}, P^{101} and R^{119}), 13 of which are conserved and 2 conservatively substituted in Dila-ERp57 (Fig. 4); 12 and 3 respectively, across species (Fig. 4); and 12 residues from the **a'** domain (K^{366}, P^{404}, W^{405}, C^{406}, G^{407}, H^{408}, K^{410}, N^{411}, T^{437}, V^{447}, R^{448}, and F^{450}) of which 10 are totally conserved, one conservatively substituted and only one not conserved throughout species (Fig. 4). All this points to an evolutionary conservation of the lectin-independent interaction between TPN and ERp57; however the corresponding cysteine residue and many of the TPN residues described to be involved in the non-covalent interactions are not conserved in non-mammalian tapasins including Dila-TPN [40], raising questions on how this interaction might occur in non-mammalian vertebrate species [40].

Two segments of low complexity were also identified along the sea bass protein (Fig. 4). A nuclear localization signal (NLS) and an ER retention KDEL-like sequence are present in the C-terminal region of Dila-ERp57 and are predominant features across species (Fig. 4). This suggests that besides putative ER localization, Dila-ERp57 may also be imported to the nucleus. Interestingly, each salmonid species has one NLS-containing and one NLS-less molecule (Fig. 4), which can be related to function differentiation, as recently reported [60]. Furthermore, consistent with its distinct basic nature, the C-terminal region has been shown to enhance ERp57 binding to CRT [69]. As this feature of the ERp57 carboxyl tail is conserved across species (Fig. 4), its specific role in CRT binding may be evolutionarily conserved as well.

3.3. 3D modelling

The crystal structures of yeast PDI (yPDI) that revealed two variations of the arrangement of its four domains, yielding a 'twisted U'-like [13] or a 'boat shaped' molecule [18], were found to be weak templates for modelling Dila-PDI mostly due to very low sequence identity (< 30%). Despite the lack of an experimental structure of full-length human PDI (hPDI), several crystal or solution structures of hPDI fragments exist, namely of the **a** [51], **b** [76], **b'x** [77], or **a'** domain (unpublished, [PDB: 1x5c]), and also of the **bb'** tandem [16], and more recently of the **bb'xa'** domain [14]. All these turned out to be suitable templates for modelling the sea bass molecule. The Dila-PDI structure was predicted by comparative modelling using the isolated hPDI **a** domain [PDB: 1mek; [51]] and bb'xa' segment [PDB: 3uem; [14]] NMR and crystal structures as templates, respectively (Fig. 5A and B). Superposing the two Dila-PDI models to the human ERp57 structure [28] enabled visualization of the full molecule. Together, the partial homology models of Dila-PDI are structurally compatible with the four Trx-like domains described for other PDI molecules (Fig. 5A). As described for the experimental structures of human and yeast PDI, each individual domain of sea bass PDI (**a**, **b**, **b'** and **a'**) assumes the typical Trx fold, i.e., a five-stranded central β-sheet structure with two α-helices on each side (Fig. 5A), and the topological arrangement βαβαβαββα (only α3-helix is missing at the **a'** domain). Furthermore, Dila-PDI also fits well the compact structural module described for the reduced hPDI bb'xa', where the **a'** domain tightly packs with the **b'** domain and x-linker [14]. Superposing the partial homology models of Dila-PDI with full-length hERp57 [28] crystal structures reveals compatibility with overall shape of the molecules. Functionally and structurally relevant residues that are conserved in Dila-PDI were mapped to both 1mek- and 3uem-based models (Fig. 5A and B).

The Dila-ERp57 structure was predicted by homology modelling using the human ERp57 [PDB entry 3f8u; [28]] crystal structure as template for the full molecule (Fig. 5C and D). Despite the fact that both **a** and **a'** domains are fixed by interactions with TPN, the crystal structure of hERp57 [28] revealed an arrangement of the four Trx-like domains highly similar to that of yPDI [13], in accordance to previous data on the shape of free ERp57 in solution [27]. The homology model of Dila-ERp57 is structurally compatible with the four Trx-like domains described for the hERp57 molecule (Fig. 5C), but minor variations to the prototypical Trx fold occur in the **b** (βαβαβαβ**α**βα) and **a'** (βαβαβ_ββα) domains, bearing an additional helix and missing helix α3, respectively. Conserved residues involved in TPN and CNX/CRT binding, as well as other structurally relevant residues were mapped (Fig. 5C and D).

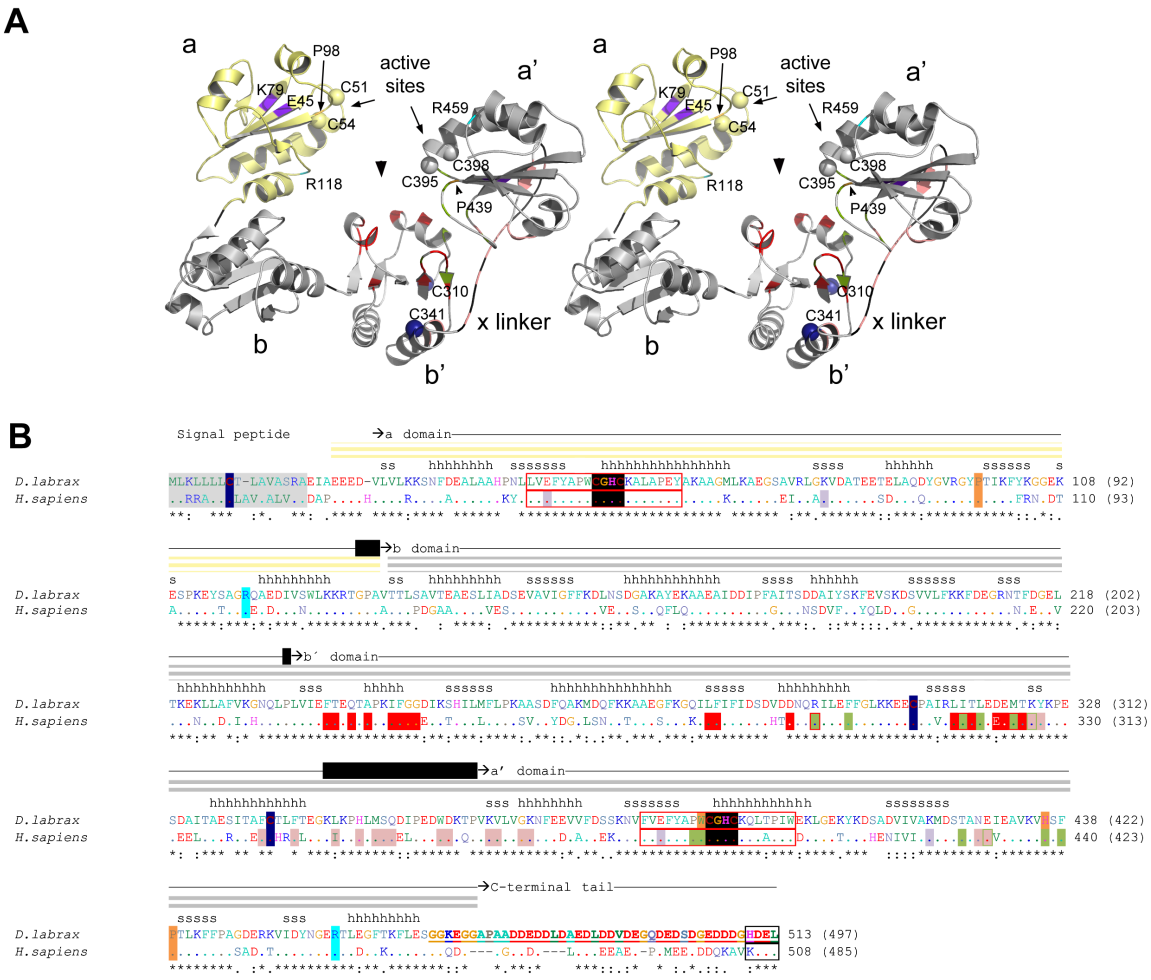
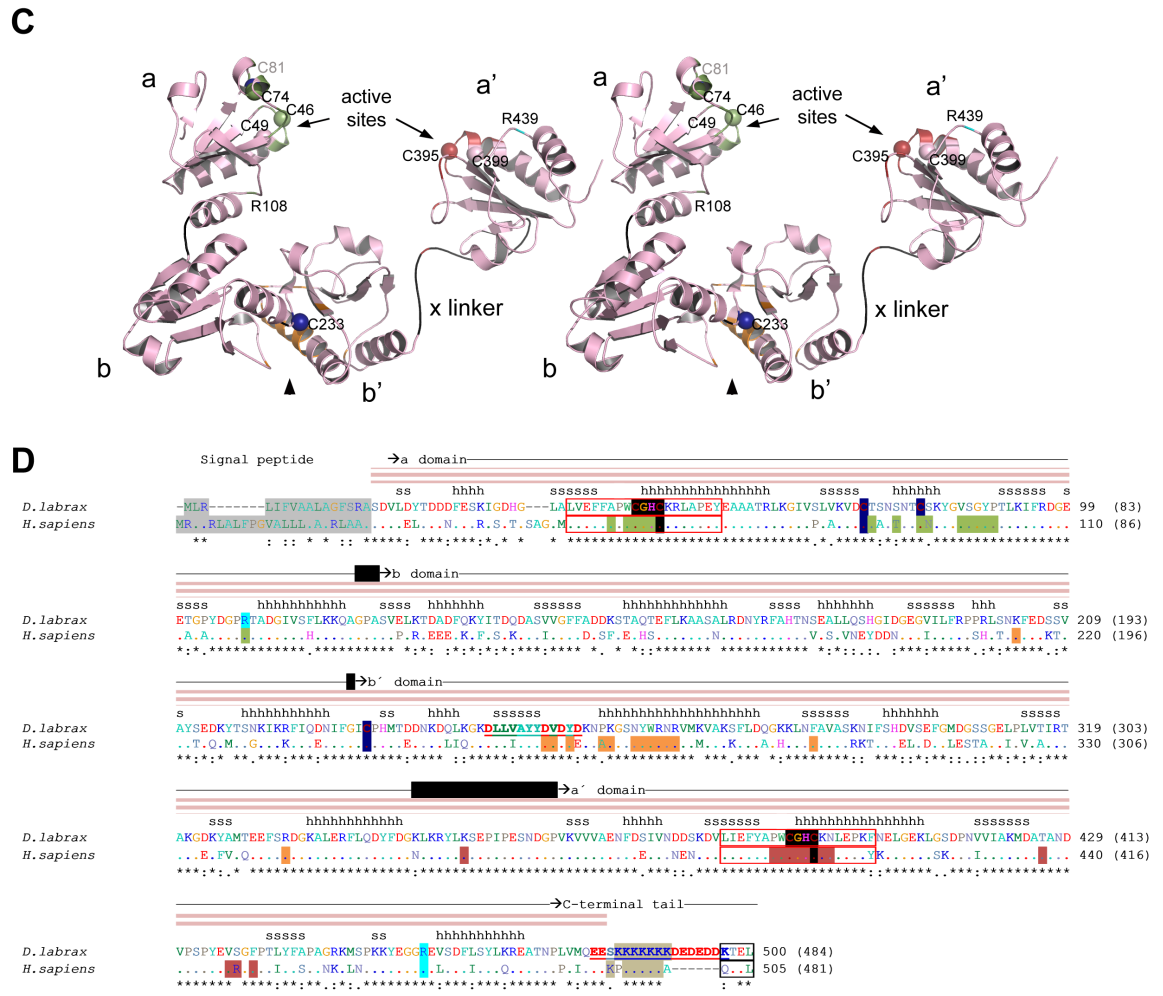


Figure 5. Three-dimensional homology models of sea bass PDI and ERp57. Dila-PDI models were generated using the human PDI **a** domain NMR structure [PDB entry 1mek; [51]] and human PDI bb'xa crystal structure [PDB entry 3uem; [14]] as templates. Dila-ERp57 model was generated using human ERp57 crystal structure [PDB entry 3f8u; [28]] as template. **A) Stereo view of Dila-PDI homology models.** The **a** domain, based on 1mek template, is colored yellow, while the bb'xa portion, based on 3uem structure, is colored grey; linker regions are in black. Superposing both partial models to human ERp57 [PDB entry 3f8u; [28]] generated the combined model. All cysteines are represented as spheres. Non-catalytic cysteines are colored dark blue. Residues possibly involved in triggering PDI conformational changes, including Cis-prolines near each active site [13], are shaded in orange [14]. Arginine residues relevant for reoxidation are colored cyan [70]. Two buried polar residues related to the redox state [13] are colored purple. Contacts of x-linker with **b'** or **a'** are colored salmon; direct contacts between **b'** and **a'** domains are colored green [14]. Residues identified as part of the substrate binding site of **b'** domain are colored red [16, 66]. The side of interaction with substrate is denoted with an arrowhead. For reference, some of the highlighted residues are labeled. **B) Alignment of target and template amino acid sequences**. The protein sequences of PDI from *Dicentrarchus labrax* and *Homo sapiens* were used in the multiple sequence alignment (see details in Figure 3). Secondary structure is indicated above the alignment with s (β strand) and h (α helix). Modeled regions are also indicated with yellow/grey horizontal bars above the alignment. **C) Stereo view of Dila-ERp57 homology model.** Dila-ERp57 model is colored light pink with linker regions in black. All cysteines are represented as spheres. Non-catalytic cysteines are in dark blue. Arginine residues relevant for reoxidation are in cyan [70]. Residues involved in TPN binding are colored green (**a** domain) and dark salmon (**a´** domain). Residues identified as



|***Figure 5*** *(continued)*| part of the calnexin/calreticulin binding site of **b'** domain are colored orange. The side of interaction with CNX/CRT P domain is denoted with an arrowhead. For reference, some of the highlighted residues are labeled. **D) Alignment of target and template amino acid sequences**. The protein sequences of ERp57 from *Dicentrarchus labrax* and *Homo sapiens* were used in the multiple sequence alignment (see details in Figure 4). Secondary structure is indicated above the alignment with s (β strand) and h (α helix). Modeled regions are indicated with pink horizontal bars above the alignment.

3.4. Phylogenetic analysis

In order to analyse the evolutionary relationships of PDI (PDIA1) and ERp57 (PDIA3) proteins from different organisms, a joint neighbour-joining tree was constructed including also representatives of PDIA2, PDIA4, PDIA5 and PDIA6 molecules (Table S3), with all groups of proteins clearly separated (Fig. 6). However, yeast PDI (PDIA1) stands outside of the displayed clusters, reflecting its ancestral existence. Within the PDI (PDIA1) group, each class of organisms segregates with high bootstrapping. As expected, Dila-PDI clusters with other bony fish (Actinopterygii) sequences being closely related to *T. nigroviridis* and *O. niloticus* PDI, the last one also from the order Perciformes. A similar arrangement is observed within the ERp57 cluster: organisms segregating by class,

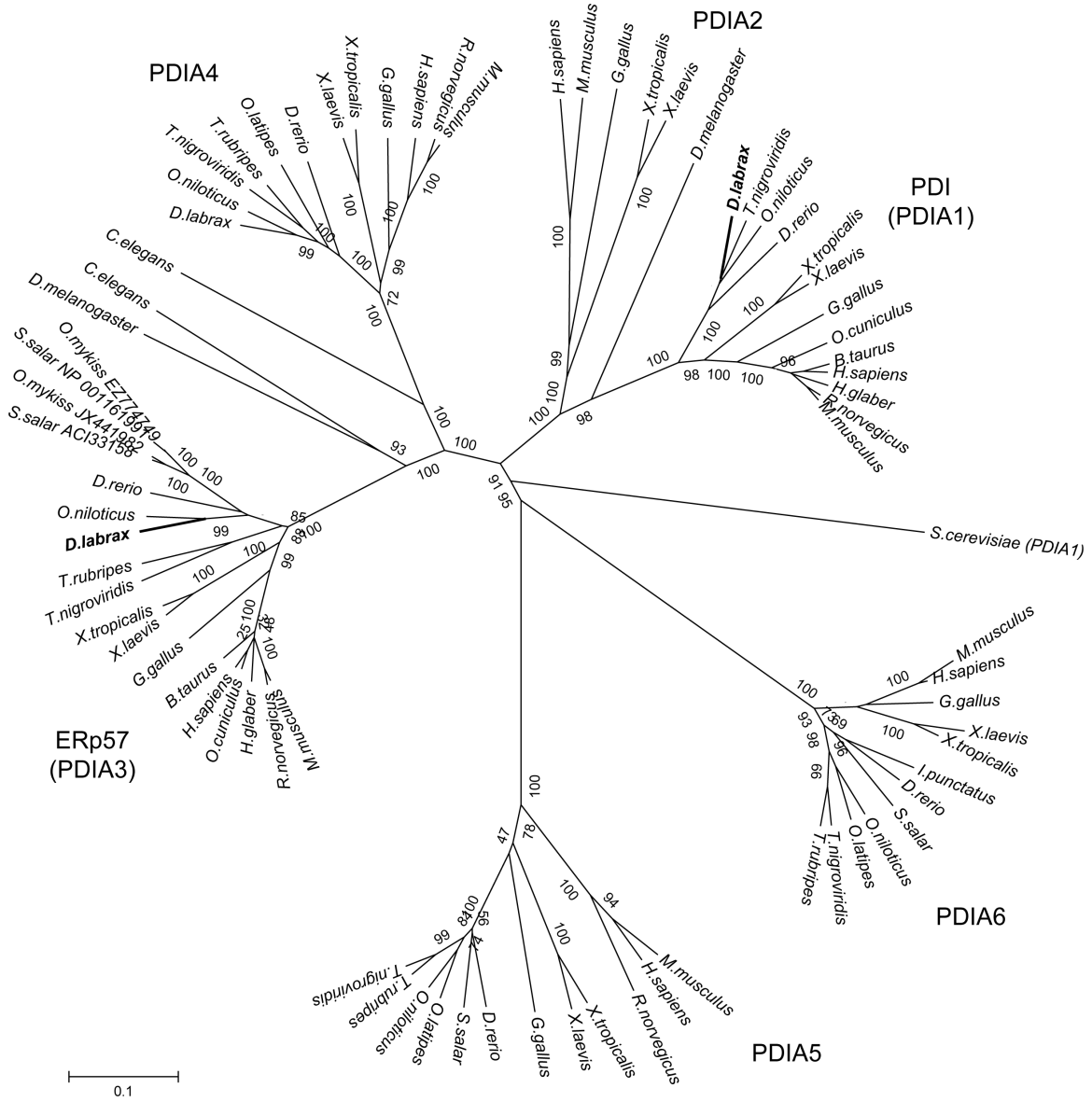


Figure 6. **Phylogeny of PDI and ERp57 and other PDI family members.** Neighbor-joining tree calculated with MEGA v5.05 [48], using p-distance parameter and pairwise deletion of gaps. The branches were validated by bootstrap analysis from 1000 replications, represented by percentages in branch nodes. Representatives from PDI (PDIA1), ERp57 (PDIA3), PDIA2 and PDIA4-6 were included in the analysis (Table S3). Sequences of precursor proteins from all molecules were used.

except for the two invertebrates that cluster together, and Dila-ERp57 being more similar to *O. niloticus*, among the teleost species. Additionally, a series of pair-wise alignments were performed with MatGAT in order to calculate similarity and identity of PDI and ERp57 amino acid sequences between species. The results confirm those previously obtained with the neighbour-joining method (Table S4). Highest identity is observed in both molecules with the mentioned *O. niloticus* (~88% and 87%, for PDI and ERp57, respectively) and the lowest with *S. cerevisiae* (~29%), followed by roundworm and fruit fly (46 and 47%, respectively). Considering the different classes of organisms analysed,

those displaying higher similarity towards European sea bass molecules are the bony fish, despite the striking degree of conservation observed across all vertebrates (>69% identity). Dila-PDI and Dila-ERp57 precursors share 34% identical and 53.6% similar amino acids.

4. Conclusions

In the present study, the European sea bass PDI (PDIA1) and ERp57 (PDIA3) transcripts and genes were successfully identified and characterized. As observed for counterparts from other species, Dila-PDI and Dila-ERp57 share the same basic domain organization with typical domains and other critical features conserved in both. Distinct properties between them include some indels, the nature of the **b'** domain and the nature and size of the C-terminal region. Other characteristics such as gene organization are also distinct between them, but conserved across species. Joint phylogenetic analysis segregated these two members of the Trx superfamily, placing sea bass molecules among the Actinopterygii clusters, with high identity observed among these enzymes in the different species. Both Dila-PDI and Dila-ERp57 3D homology models have been predicted and compared to human homologs. Altogether these results point to an orthology between the reported molecules and those described for other species. Hence, important primary data has been provided to perform further studies on the diverse PDI and ERp57 functions in immunity, specifically their involvement in MHC class I antigen presentation in this species.

Acknowledgements

FEDER funds through the Operational Competitiveness Programme – COMPETE and national funds through FCT – Fundação para a Ciência e a Tecnologia under projects FCOMP-01-0124-FEDER-022718 (PEst-C/SAU/LA0002/2011) and FCOMP-01-0124-FEDER-007173 (POCTI/CVT/44925/2006) supported this work. Rute D. Pinto acknowledges FCT PhD fellowship BD/42327/2007, financed by POPH-QREN and co-funded by FSE and MCTES. The funders had no role in study design, data collection and analysis, decision to publish, or publication of the manuscript. No additional external funding was received for this study.

References

1. Galligan JJ, Petersen DR. The human protein disulfide isomerase gene family. Hum Genomics. 2012 6:6.
2. Maattanen P, Gehring K, Bergeron JJ, Thomas DY. Protein quality control in the ER: the recognition of misfolded proteins. Semin Cell Dev Biol. 2010 21:500-11.
3. Appenzeller-Herzog C, Ellgaard L. The human PDI family: versatility packed into a single fold. Biochim Biophys Acta. 2008 1783:535-48.
4. Kozlov G, Maattanen P, Thomas DY, Gehring K. A structural overview of the PDI family of proteins. FEBS J. 2010 277:3924-36.
5. Collet JF, Messens J. Structure, function, and mechanism of thioredoxin proteins. Antioxid Redox Signal. 2010 13:1205-16.
6. Kozlov G, Pocanschi CL, Rosenauer A, Bastos-Aristizabal S, Gorelik A, Williams DB, et al. Structural basis of carbohydrate recognition by calreticulin. J Biol Chem. 2010 285:38612-20.
7. Givol D, Goldberger RF, Anfinsen CB. Oxidation and Disulfide Interchange in the Reactivation of Reduced Ribonuclease. J Biol Chem. 1964 239:PC3114-16.
8. Hatahet F, Ruddock LW. Protein disulfide isomerase: a critical evaluation of its function in disulfide bond formation. Antioxid Redox Signal. 2009 11:2807-50.
9. Inaba K. Structural basis of protein disulfide bond generation in the cell. Genes Cells. 2010 15:935-43.
10. Cai H, Wang CC, Tsou CL. Chaperone-like activity of protein disulfide isomerase in the refolding of a protein with no disulfide bonds. J Biol Chem. 1994 269:24550-2.
11. Laurindo FR, Pescatore LA, Fernandes Dde C. Protein disulfide isomerase in redox cell signaling and homeostasis. Free Radic Biol Med. 2012 52:1954-69.
12. Freedman RB, Gane PJ, Hawkins HC, Hlodan R, McLaughlin SH, Parry JW. Experimental and theoretical analyses of the domain architecture of mammalian protein disulphide-isomerase. Biol Chem. 1998 379:321-8.
13. Tian G, Xiang S, Noiva R, Lennarz WJ, Schindelin H. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. Cell. 2006 124:61-73.
14. Wang C, Yu J, Huo L, Wang L, Feng W, Wang CC. Human protein-disulfide isomerase is a redox-regulated chaperone activated by oxidation of domain a'. J Biol Chem. 2012 287:1139-49.
15. Walker KW, Gilbert HF. Scanning and escape during protein-disulfide isomerase-assisted protein folding. J Biol Chem. 1997 272:8845-8.

16. Denisov AY, Maattanen P, Dabrowski C, Kozlov G, Thomas DY, Gehring K. Solution structure of the bb' domains of human protein disulfide isomerase. FEBS J. 2009 276:1440-9.
17. Pirneskoski A, Klappa P, Lobell M, Williamson RA, Byrne L, Alanen HI, et al. Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. J Biol Chem. 2004 279:10374-81.
18. Tian G, Kober FX, Lewandrowski U, Sickmann A, Lennarz WJ, Schindelin H. The catalytic activity of protein-disulfide isomerase requires a conformationally flexible molecule. J Biol Chem. 2008 283:33630-40.
19. Bennett CF, Balcarek JM, Varrichio A, Crooke ST. Molecular cloning and complete amino-acid sequence of form-I phosphoinositide-specific phospholipase C. Nature. 1988 334:268-70.
20. Coe H, Michalak M. ERp57, a multifunctional endoplasmic reticulum resident oxidoreductase. Int J Biochem Cell Biol. 2010 42:796-9.
21. Coppari S, Altieri F, Ferraro A, Chichiarelli S, Eufemi M, Turano C. Nuclear localization and DNA interaction of protein disulfide isomerase ERp57 in mammalian cells. J Cell Biochem. 2002 85:325-33.
22. Guo GG, Patel K, Kumar V, Shah M, Fried VA, Etlinger JD, et al. Association of the chaperone glucose-regulated protein 58 (GRP58/ER-60/ERp57) with Stat3 in cytosol and plasma membrane complexes. J Interferon Cytokine Res. 2002 22:555-63.
23. Panaretakis T, Joza N, Modjtahedi N, Tesniere A, Vitale I, Durchschlag M, et al. The co-translocation of ERp57 and calreticulin determines the immunogenicity of cell death. Cell Death Differ. 2008 15:1499-509.
24. Khanal RC, Nemere I. The ERp57/GRp58/1,25D3-MARRS receptor: multiple functional roles in diverse cell systems. Curr Med Chem. 2007 14:1087-93.
25. Jordan PA, Gibbins JM. Extracellular disulfide exchange and the regulation of cellular function. Antioxid Redox Signal. 2006 8:312-24.
26. Holbrook LM, Sasikumar P, Stanley RG, Simmonds AD, Bicknell AB, Gibbins JM. The platelet-surface thiol isomerase enzyme ERp57 modulates platelet function. J Thromb Haemost. 2012 10:278-88.
27. Kozlov G, Maattanen P, Schrag JD, Pollock S, Cygler M, Nagar B, et al. Crystal structure of the bb' domains of the protein disulfide isomerase ERp57. Structure. 2006 14:1331-9.
28. Dong G, Wearsch PA, Peaper DR, Cresswell P, Reinisch KM. Insights into MHC class I peptide loading from the structure of the tapasin-ERp57 thiol oxidoreductase

heterodimer. Immunity. 2009 30:21-32.

29. Maattanen P, Kozlov G, Gehring K, Thomas DY. ERp57 and PDI: multifunctional protein disulfide isomerases with similar domain architectures but differing substrate-partner associations. Biochem Cell Biol. 2006 84:881-9.
30. Frickel EM, Frei P, Bouvier M, Stafford WF, Helenius A, Glockshuber R, et al. ERp57 is a multifunctional thiol-disulfide oxidoreductase. J Biol Chem. 2004 279:18277-87.
31. Williams DB. Beyond lectins: the calnexin/calreticulin chaperone system of the endoplasmic reticulum. J Cell Sci. 2006 119:615-23.
32. Zapun A, Darby NJ, Tessier DC, Michalak M, Bergeron JJ, Thomas DY. Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. J Biol Chem. 1998 273:6009-12.
33. Wearsch PA, Cresswell P. The quality control of MHC class I peptide loading. Curr Opin Cell Biol. 2008 20:624-31.
34. Zhang Y, Baig E, Williams DB. Functions of ERp57 in the folding and assembly of major histocompatibility complex class I molecules. J Biol Chem. 2006 281:14622-31.
35. Peaper DR, Wearsch PA, Cresswell P. Tapasin and ERp57 form a stable disulfide-linked dimer within the MHC class I peptide-loading complex. EMBO J. 2005 24:3613-23.
36. Radcliffe CM, Diedrich G, Harvey DJ, Dwek RA, Cresswell P, Rudd PM. Identification of specific glycoforms of major histocompatibility complex class I heavy chains suggests that class I peptide loading is an adaptation of the quality control pathway involving calreticulin and ERp57. J Biol Chem. 2002 277:46415-23.
37. Park B, Lee S, Kim E, Cho K, Riddell SR, Cho S, et al. Redox regulation facilitates optimal peptide selection by MHC class I during antigen processing. Cell. 2006 127:369-82.
38. Cho K, Cho S, Lee SO, Oh C, Kang K, Ryoo J, et al. Redox-regulated peptide transfer from the transporter associated with antigen processing to major histocompatibility complex class I molecules by protein disulfide isomerase. Antioxid Redox Signal. 2011 15:621-33.
39. Lee S, Park B, Kang K, Ahn K. Redox-regulated export of the major histocompatibility complex class I-peptide complexes from the endoplasmic reticulum. Mol Biol Cell. 2009 20:3285-94.
40. Pinto RD, da Silva DV, Pereira PJ, dos Santos NM. Molecular cloning and characterization of sea bass (Dicentrarchus labrax, L.) Tapasin. Fish Shellfish Immunol. 2012 32:110-20.

41. Pinto RD, Pereira PJ, dos Santos NM. Transporters associated with antigen processing (TAP) in sea bass (Dicentrarchus labrax, L.): molecular cloning and characterization of TAP1 and TAP2. Dev Comp Immunol. 2011 35:1173-81.
42. Pinto RD, Randelli E, Buonocore F, Pereira PJ, dos Santos NM. Molecular cloning and characterization of sea bass (Dicentrarchus labrax, L.) MHC class I heavy chain and beta2-microglobulin. Dev Comp Immunol. 2013 39:234-54.
43. Pinto RD, Moreira AR, Pereira PJ, Dos Santos NM. Molecular cloning and characterization of sea bass (Dicentrarchus labrax, L.) calreticulin. Fish Shellfish Immunol. 2013.
44. Stet RJ, van Erp SH, Hermsen T, Sultmann HA, Egberts E. Polymorphism and estimation of the number of MhcCyca class I and class II genes in laboratory strains of the common carp (Cyprinus carpio L.). Dev Comp Immunol. 1993 17:141-56.
45. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007 23:2947-8.
46. Hall T. BioEdit. Biological sequence alignment editor for Windows. North Carolina State University, NC, USA; ; 1998.
47. Bairoch A, Bucher P, Hofmann K. The PROSITE database, its status in 1997. Nucleic Acids Res. 1997 25:217-21.
48. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011 28:2731-9.
49. Campanella JJ, Bitincka L, Smalley J. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. BMC Bioinformatics. 2003 4:29.
50. F.M. Ausubel RB, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl Current Protocols in Molecular Biology. New York John Wiley and Sons, Inc; 1999.
51. Kemmink J, Darby NJ, Dijkstra K, Nilges M, Creighton TE. Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear 13C/15N NMR spectroscopy. Biochemistry. 1996 35:7684-91.
52. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006 22:195-201.
53. Hutchinson EG, Thornton JM. PROMOTIF--a program to identify and analyze structural motifs in proteins. Protein Sci. 1996 5:212-20.

54. Hooft RW, Vriend G, Sander C, Abola EE. Errors in protein structures. Nature. 1996 381:272.
55. Laskowski R A MMW, Moss D, Thornton J M PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst. 1993 26:283-91.
56. Shaw G, Kamen R. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell. 1986 46:659-67.
57. Tasanen K, Parkkonen T, Chow LT, Kivirikko KI, Pihlajaniemi T. Characterization of the human gene for a polypeptide that acts both as the beta subunit of prolyl 4-hydroxylase and as protein disulfide isomerase. J Biol Chem. 1988 263:16218-24.
58. Koivunen P, Horelli-Kuitunen N, Helaakoski T, Karvonen P, Jaakkola M, Palotie A, et al. Structures of the human gene for the protein disulfide isomerase-related polypeptide ERp60 and a processed gene and assignment of these genes to 15q15 and 1q21. Genomics. 1997 42:397-404.
59. Huang TS, Olsvik PA, Krovel A, Tung HS, Torstensen BE. Stress-induced expression of protein disulfide isomerase associated 3 (PDIA3) in Atlantic salmon (Salmo salar L.). Comp Biochem Physiol B Biochem Mol Biol. 2009 154:435-42.
60. Sever L, Bols NC, Dixon B. The cloning and inducible expression of the rainbow trout ERp57 gene. Fish Shellfish Immunol. 2012.
61. Xiao R, Wilkinson B, Solovyov A, Winther JR, Holmgren A, Lundstrom-Ljung J, et al. The contributions of protein disulfide isomerase and its homologues to oxidative protein folding in the yeast endoplasmic reticulum. J Biol Chem. 2004 279:49780-6.
62. Wilkinson B, Xiao R, Gilbert HF. A structural disulfide of yeast protein-disulfide isomerase destabilizes the active site disulfide of the N-terminal thioredoxin domain. J Biol Chem. 2005 280:11483-7.
63. Schwaller M, Wilkinson B, Gilbert HF. Reduction-reoxidation cycles contribute to catalysis of disulfide isomerization by protein-disulfide isomerase. J Biol Chem. 2003 278:7154-9.
64. Darby NJ, Creighton TE. Characterization of the active site cysteine residues of the thioredoxin-like domains of protein disulfide isomerase. Biochemistry. 1995 34:16770-80.
65. Klappa P, Ruddock LW, Darby NJ, Freedman RB. The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. Embo J. 1998 17:927-35.
66. Byrne LJ, Sidhu A, Wallis AK, Ruddock LW, Freedman RB, Howard MJ, et al. Mapping of the ligand-binding site on the b' domain of human PDI: interaction with peptide ligands and the x-linker region. Biochem J. 2009 423:209-17.

67. Grillo C, D'Ambrosio C, Consalvi V, Chiaraluce R, Scaloni A, Maceroni M, et al. DNA-binding activity of the ERp57 C-terminal domain is related to a redox-dependent conformational change. J Biol Chem. 2007 282:10299-310.
68. Oliver JD, Roderick HL, Llewellyn DH, High S. ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. Mol Biol Cell. 1999 10:2573-82.
69. Russell SJ, Ruddock LW, Salo KE, Oliver JD, Roebuck QP, Llewellyn DH, et al. The primary substrate binding site in the b' domain of ERp57 is adapted for endoplasmic reticulum lectin association. J Biol Chem. 2004 279:18861-9.
70. Lappi AK, Lensink MF, Alanen HI, Salo KE, Lobell M, Juffer AH, et al. A conserved arginine plays a role in the catalytic cycle of the protein disulphide isomerases. J Mol Biol. 2004 335:283-95.
71. Lindquist JA, Hammerling GJ, Trowsdale J. ER60/ERp57 forms disulfide-bonded intermediates with MHC class I heavy chain. FASEB J. 2001 15:1448-50.
72. Antoniou AN, Ford S, Alphey M, Osborne A, Elliott T, Powis SJ. The oxidoreductase ERp57 efficiently reduces partially folded in preference to fully folded MHC class I molecules. Embo J. 2002 21:2655-63.
73. Zhang Y, Kozlov G, Pocanschi CL, Brockmeier U, Ireland BS, Maattanen P, et al. ERp57 does not require interactions with calnexin and calreticulin to promote assembly of class I histocompatibility molecules, and it enhances peptide loading independently of its redox activity. J Biol Chem. 2009 284:10160-73.
74. Wearsch PA, Peaper DR, Cresswell P. Essential glycan-dependent interactions optimize MHC class I peptide loading. Proc Natl Acad Sci U S A. 2011 108:4950-5.
75. Dick TP, Bangia N, Peaper DR, Cresswell P. Disulfide bond isomerization and the assembly of MHC class I-peptide complexes. Immunity. 2002 16:87-98.
76. Kemmink J, Dijkstra K, Mariani M, Scheek RM, Penka E, Nilges M, et al. The structure in solution of the b domain of protein disulfide isomerase. J Biomol NMR. 1999 13:357-68.
77. Nguyen VD, Wallis K, Howard MJ, Haapalainen AM, Salo KE, Saaranen MJ, et al. Alternative conformations of the x region of human protein disulphide-isomerase modulate exposure of the substrate binding b' domain. J Mol Biol. 2008 383:1144-55.

Supplementary data

Table S1

Oligonucleotide sequences.

Designation	Nucleotide sequence 5'- 3'	Use / Function
APv	GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTTTV	cDNA synthesis/5'RACE/full cDNA id
AUAP	GGCCACGCGTCGACTAGTAC	5'RACE/full cDNA id
APv2	GACTCAGGACTTCAGGACTTAGTTTTTTTTTTTTTTTTTV	5'RACE
AUAP2	GACTCAGGACTTCAGGACTTAG	5'RACE
T7	TAATACGACTCACTATAGGGCGA	sequencing
SP6	CTATTTAGGTGACACTATAGAATAC	sequencing
DLPDIFW2	GTGCGATGCCACGTCAGTAG	full cDNA & partial gene id/sequencing
DLPDIFW3	AACCGTAGCCGACCGTGTCT	full cDNA & partial gene id
DLPDIFW4	AAGCAGCTGAAGCCATAGAC	sequencing
DLPDIFW5	TGCACCTTGGTGTGGACACT	partial gene id
DLPDIFW6	CATTCCCTTCGCCATCACAT	partial gene id
DLPDIFW7	CAGCGCATCCTGGAGTTCTT	partial gene id
DLPDIFW10	CTGCCTCTGGTCATCGAGTT	partial gene id
DLPDIFW11	GATTACAATGGGGAGAGAAC	sequencing
DLPDIFW12	TGAGGGGCTGTGTAGGTCAT	sequencing
DLPDIFW13	GAACATGCCTCCCCATATCA	sequencing
DLPDIFW14	GGATGCATGTTGAGCTTTGAC	sequencing
DLPDIFW15	GAGAAGTCGGCCACGATCAC	sequencing
DLPDIFW16	CTGGCCACCTTTTCCTTCTT	sequencing
DLPDIFW17	GGGTAATGAGCGTACAGTTG	sequencing
DLPDIFW18	GATATGTAATCCTCCAAGTTCTC	sequencing
DLPDIFW19	GAAGTTGGAGGCGGTAGTGG	sequencing
DLPDIRV2	ATGCAGCTTTGGGCAGAAAC	5'RACE (cDNA synthesis)/partial gene id
DLPDIRV3	AGATTTTAGGGGCGGTCTGC	5'RACE/sequencing
DLPDIRV4	GCAGCTGGTTGCCCTTGAC	5'RACE
DLPDIRV5	CCTCCTTACCGCCGCTCT	sequencing
DLPDIRV6	GGAGGAGGGGAGCCAATTC	partial gene id/sequencing
DLPDIRV8	ACTTAAGAGTGGGGAAGCTG	partial gene id
DLPDIRV9	ACAACGCTGTCCTTGGACAC	partial gene id
DLPDIRV10	GTGTGCAGAATGCGGTGATG	partiall gene id
DLPDIRV13	GACAAGCATGCAGCTCTAGT	sequencing
DLPDIRV14	GTATCACGCCACACAGGTAA	sequencing
ERp57FW1	GAGTTYTTCGCNCCITGGTG	partial cDNA id
ERp57RV1	CCWCCYTCRTAYTTCTTTGG	partial cDNA id
ERp57RV2	TCATIACYCKRTTYCKCCA	partial cDNA id
DLERp57FW7	CTCCCTGGTGCGGACACT	sequencing
DLERp57FW8	TATCGACCTGTCTGCATCTG	full cDNA & gene id/southern
DLERp57FW9	GAGCTGCCACAGACACAAC	full gene id
DLERp57FW12	ACCACGACTCAGCAACAAG	sequencing
DLERp57FW13	GCACATCCAACAGCAACACA	sequencing
DLERp57FW14	GGAATCTGCCCCCACATGAC	sequencing
DLERp57FW16	CTGCTCAAAACATCCCTTAG	sequencing
DLERp57FW17	TTCAGTCGGTGACTCCAGTT	sequencing
DLERp57FW18	AGCAGTCCGATCTGTTGAAG	sequencing
DLERp57RV3	GGTAGCCACTGACACCATA	5'RACE
DLERp57RV4	ATAGGGACCAGTTTCCTCTC	5'RACE
DLERp57RV5	ACAATCCCATCTGCAGTCCT	5'RACE (cDNA snthesis)
DLERp57RV6	CGAGTCCGTGGTCTCCGATCT	5'RACE/southern
DLERp57RV7	GACCAGAGAGACGATGCCTTTCAG	5'RACE/sequencing
DLERp57RV10	ATCAACATCATAATAAGCCACCAG	sequencing
DLERp57RV11	GATTTTGTGGGCCTTTTTAT	full gene id
DLERp57RV12	TCACAATTCTTCCAAATGTA	full gene id

Letter nucleotide code: V= A, C, G; I= deoxyinosine; R= A, G; M= A, C; Y= C, T; W= A, T; K= G, T; S= G, C.

Table S2

Distinct regions encoded by each exon.

	Exon	Region(s)
PDI	I	5' UTR, signal peptide, a domain part 1 (a domain-p1)
	II	a domain-p2
	III	a domain-p3, b domain-p1
	IV	b domain-p2
	V	b domain-p3, b' domain-p1
	VI	b' domain-p2
	VII	b' domain-p3, x linker-p1
	VIII	x linker-p2, a' domain-p1
	IX	a' domain-p2
	X	a' domain-p3, c-terminal tail (ctail)-p1
	XI	ctail-p2, 3' UTR
ERp57	I	5' UTR, signal peptide, a domain-p1
	II	a domain-p2
	III	a domain-p3
	IV	a domain-p4, b domain-p1
	V	b domain-p2
	VI	b domain-p3
	VII	b domain-p4, b'domain-p1
	VIII	b' domain-p2
	IX	b' domain-p3, x linker, a´domain-p1
	X	a' domain-p2
	XI	a' domain-p3
	XII	a' domain-p4
	XIII	c terminal tail, 3' UTR

Table S3

Accession numbers of sequences used in this study.

Species	PDI (PDIA1)	ERp57 (PDIA3)	PDIA2	PDIA4	PDIA5	PDIA6
D. labrax			-	CBN81869	-	XP_003458276
O. niloticus	XP_003448052	XP_003455475	-	XP_003443545	XP_003454923	-
O. latipes	-	-	-	XP_004081050	XP_004081476	XP_004083501
T. nigroviridis	CAG06136	CAG01048	-	CAF97973	CAG08268	CAG03659
T. rubripes	-	XP_003967256 ACI33158;	-	XP_003967904	XP_003962200	XP_003971872
S. salar	-	NP_001161991	-	-	NP_001133435	ACI33422
O. mykiss	-	JX441982; EZ774749	-	-	-	-
D. rerio	NP_998529	NP_001186666	-	AAI17630	AAI54607	AF387900
I. punctatus	-	-	-	-	-	AFL70280
X. laevis	AAH46736	NP_001080051	NP_001083648	NP_001088331	AAH76861	NP_001086643
X. tropicalis	AAH64877	NP_989329	NP_001011281	XP_002941424	CAJ81340	XP_002940166
G. gallus	P09102	NP_989441	XP_003642202	CAG31923	XP_422097	XP_419952
R. norvegicus	P04785	NP_059015	-	EDL88288	XP_003750011	-
M. musculus	BAE40274	NP_031978	NP_001074539	BAE27045	NP_082571	AAH06865
H. glaber	EHB09338	EHB10367	-	-	-	-
O. cuniculus	P21195	NP_001164786	-	-	-	-
B. taurus	P05307	NP_776758	-	-	-	-
H. sapiens	NP_000909	NP_005304	AAH75029	BAF83660	NP_006801	NP_005733
D. melanogaster	AAF49659	AAF58609	-	-	-	-
S. cerevisiae	NP_009887	-	-	-	-	-
C. elegans	-	NM_059594	-	P34329	-	-

Table S4

Percentages of identity and similarity calculated with MatGat.

		Similarity (%)	Identity (%)
PDI	*D. labrax*	-	-
	O. niloticus	93.8	88.0
	T. nigroviridis	89.5	84.8
	D. rerio	92.0	81.5
	X. tropicalis	84.0	69.5
	X. laevis	84.8	71.1
	G. gallus	83.1	69.4
	O. cuniculus	83.6	72.1
	R. norvegicus	83.2	72.9
	M. musculus	83.6	73.1
	H. glaber	84.4	72.5
	B. taurus	83.2	71.8
	H. sapiens	84.2	72.5
	S. cerevisiae	52.7	28.8
ERp57	*D. labrax*	-	-
	O. niloticus	94.2	87.0
	S. salar	89.6	79.8
	D. rerio	91.0	82.2
	T. nigroviridis	86.6	74.4
	X. laevis	85.1	70.3
	X. tropicalis	86.0	73.0
	G. gallus	84.8	74.1
	H. glaber	85.9	72.8
	O. cuniculus	86.5	74.2
	B. taurus	86.7	74.0
	M. musculus	85.9	71.5
	R. norvegicus	85.7	70.9
	H. sapiens	86.7	73.6
	D. melanogaster	64.6	47.0

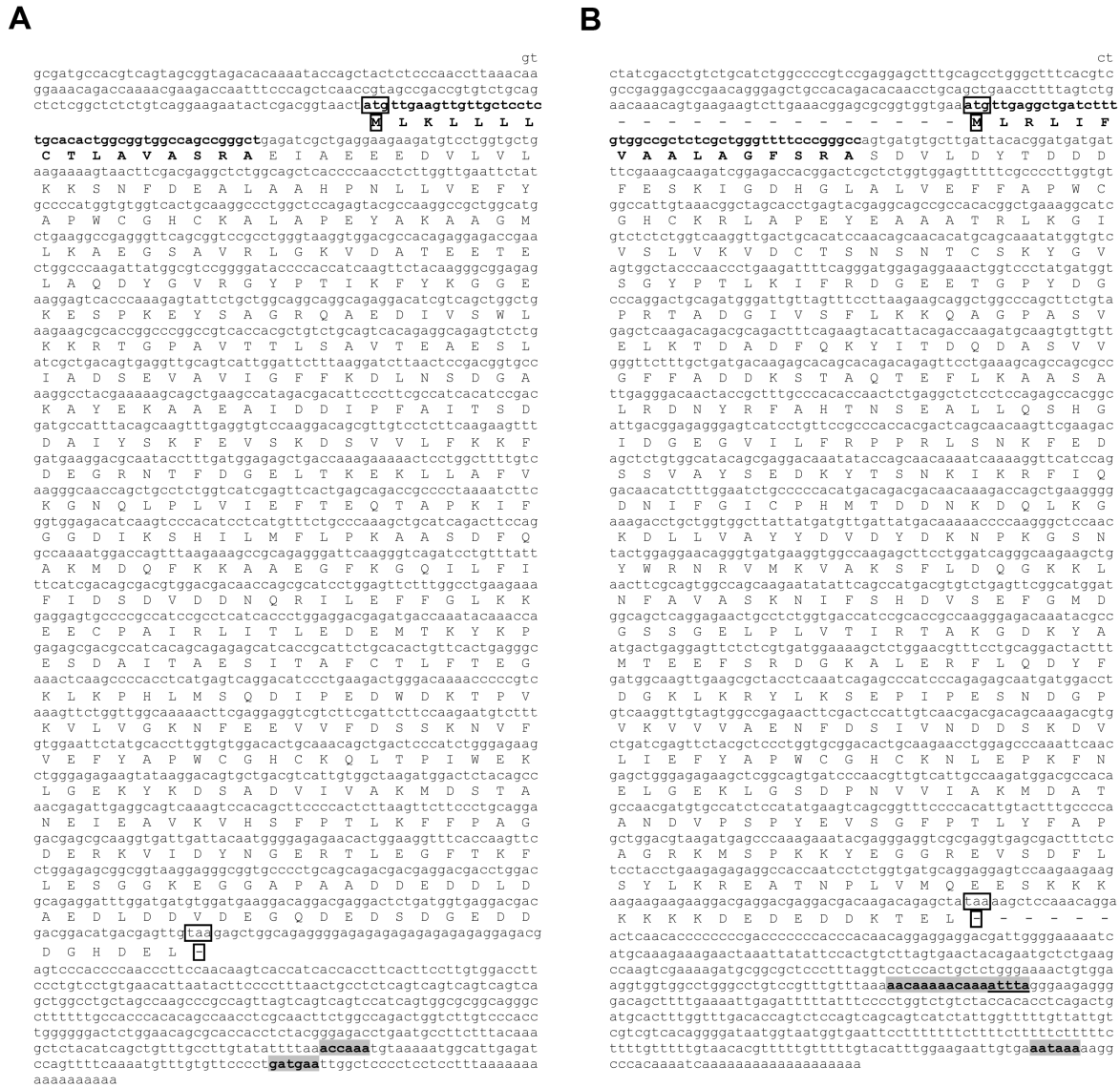


Figure S1. **Nucleotide and deduced amino acid sequences of sea bass PDI and ERp57 cDNAs.** Nucleotide sequences (untranslated regions and ORF) of sea bass PDI (A) and ERp57 (B) are shown in lowercase letters while the predicted protein sequences are shown in uppercase letters. The translation start and stop codons and the initial methionine are boxed. The predicted signal peptide is highlighted in bold type. Within the 3'UTR, the putative instability motifs are underlined while poly-adenylation signal sequences are shaded in grey and in bold type.

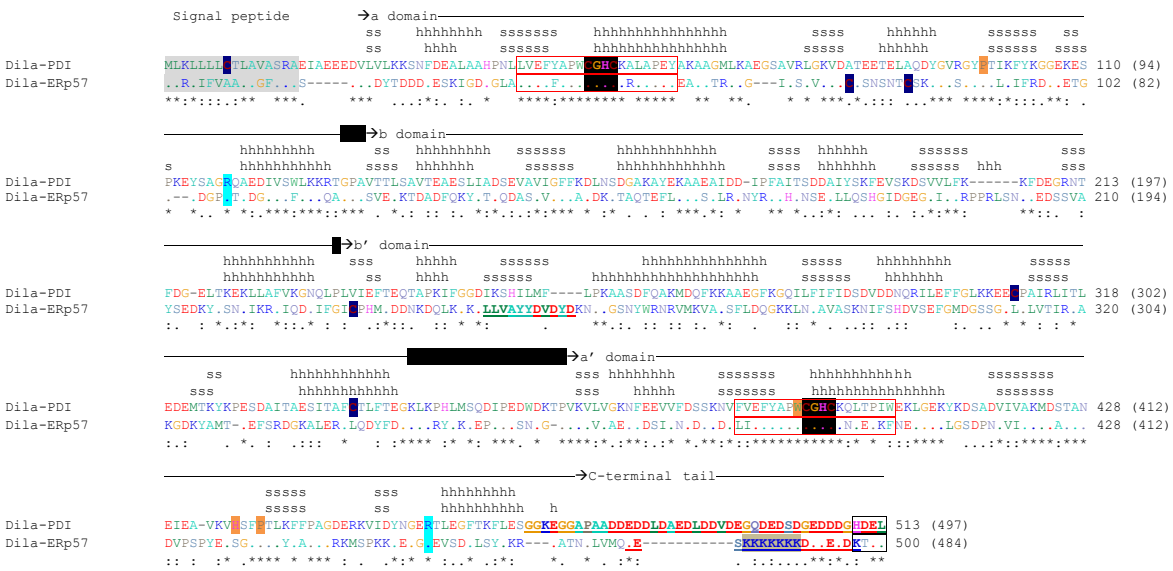


Figure S2. Joint alignment of sea bass PDI and ERp57 amino acid sequences. Amino acid sequences were retrieved from GenBank (Table S3) and aligned with CLUSTALW [45]. Dashes indicate gaps that maximize the alignment and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with (*), (:) and (.), respectively. Residue numbers are given to the right of the sequences, with numbers in parenthesis referring to the mature proteins. Signal peptides are shaded in grey and denoted above the alignment. Predicted **a**, **b**, **b'**, and **a** domains based on human PDI [14], are indicated above the alignment, with junction regions represented by black boxes. Also above the alignment are indicated the secondary structure elements from each molecule ('s' for strands and 'h' for helices). The two Trx signature motifs of domains **a** and **a'** are boxed, with each active site shaded in black. Non-catalytic cysteines are shaded in dark blue. Residues possibly involved in triggering PDI conformational changes, including cis-prolines near each active site [13], are shaded in orange [14]. Arginine residues relevant for reoxidation are shaded cyan [70]. Low complexity regions are underlined and in bold type. Putative nuclear localization signal is shaded in sepia. KDEL-like ER-retention signals are boxed.

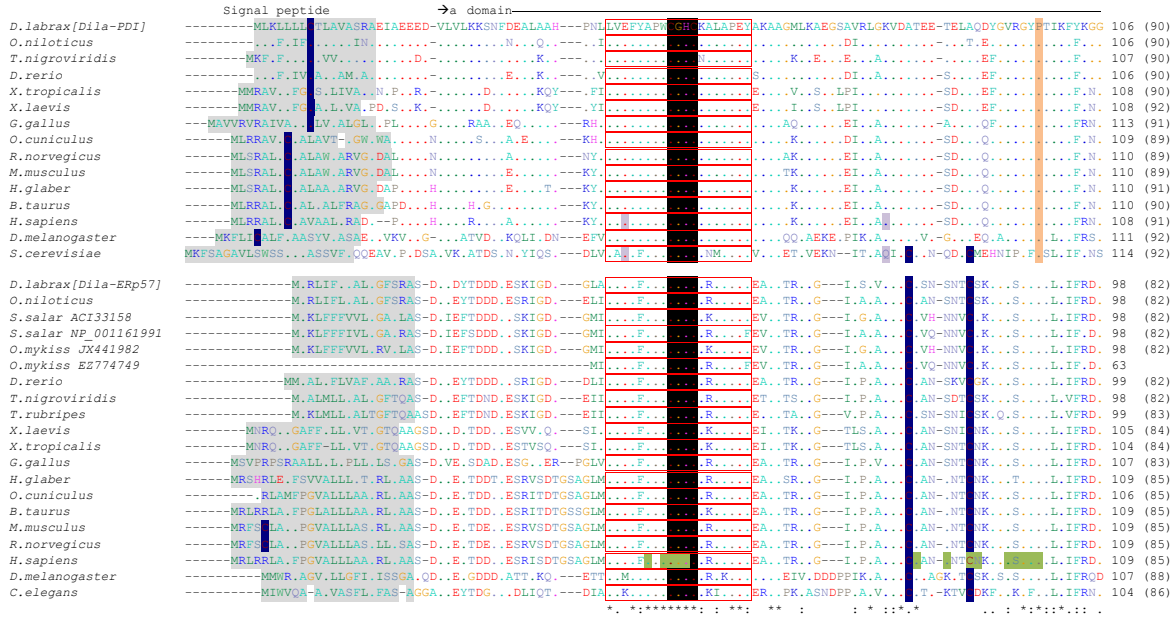


Figure S3. Joint alignment of PDI and ERp57 amino acid sequences from several species. Amino acid sequences were retrieved from GenBank (Table S3) and aligned with CLUSTALW [45]. Dashes indicate gaps that maximize the alignment and dots denote residues identical to the first sequence of the alignment. Identical residues, conserved and semi-conserved substitutions are denoted below the alignment with (*), (:) and (.), respectively. Residue numbers are given to the right of the sequences, with numbers in parenthesis referring to the mature proteins. Signal peptides are shaded in grey and denoted above the alignment. Predicted **a**, **b**, **b'**, and **a** domains, based on human PDI [14], are indicated above the alignment, with junction regions represented by black boxes. The two Trx signature motifs of domains **a** and **a'** are boxed, with each active site shaded in black. Non-catalytic cysteines are shaded in dark blue. Arginine residues relevant for reoxidation are shaded cyan [70]. Residues possibly involved in triggering PDI conformational changes, including cis-prolines near each active site [13], are shaded in orange [14]. Two buried polar residues related to the redox state [13] are shaded purple in human and yeast PDI. Hydrophobic patch on the b' domain putatively involved in peptide/substrate binding is shaded yellow in yPDI [13]. Contacts of x-linker with **b'** or **a'** are shaded in light salmon; direct contacts between **b'** and **a'** domains are shaded in green [14]. Residues identified as part of the substrate binding site of **b'** domain are shaded in red [16, 66]. Residues of human ERp57 that contact CNX/CRT are shaded orange. Residues involved in TPN binding are shaded green (**a** domain) and dark salmon (**a´** domain). Low complexity regions are underlined and in bold type. ERp57 putative nuclear localization signal is shaded in sepia. KDEL-like ER-retention signals are boxed.

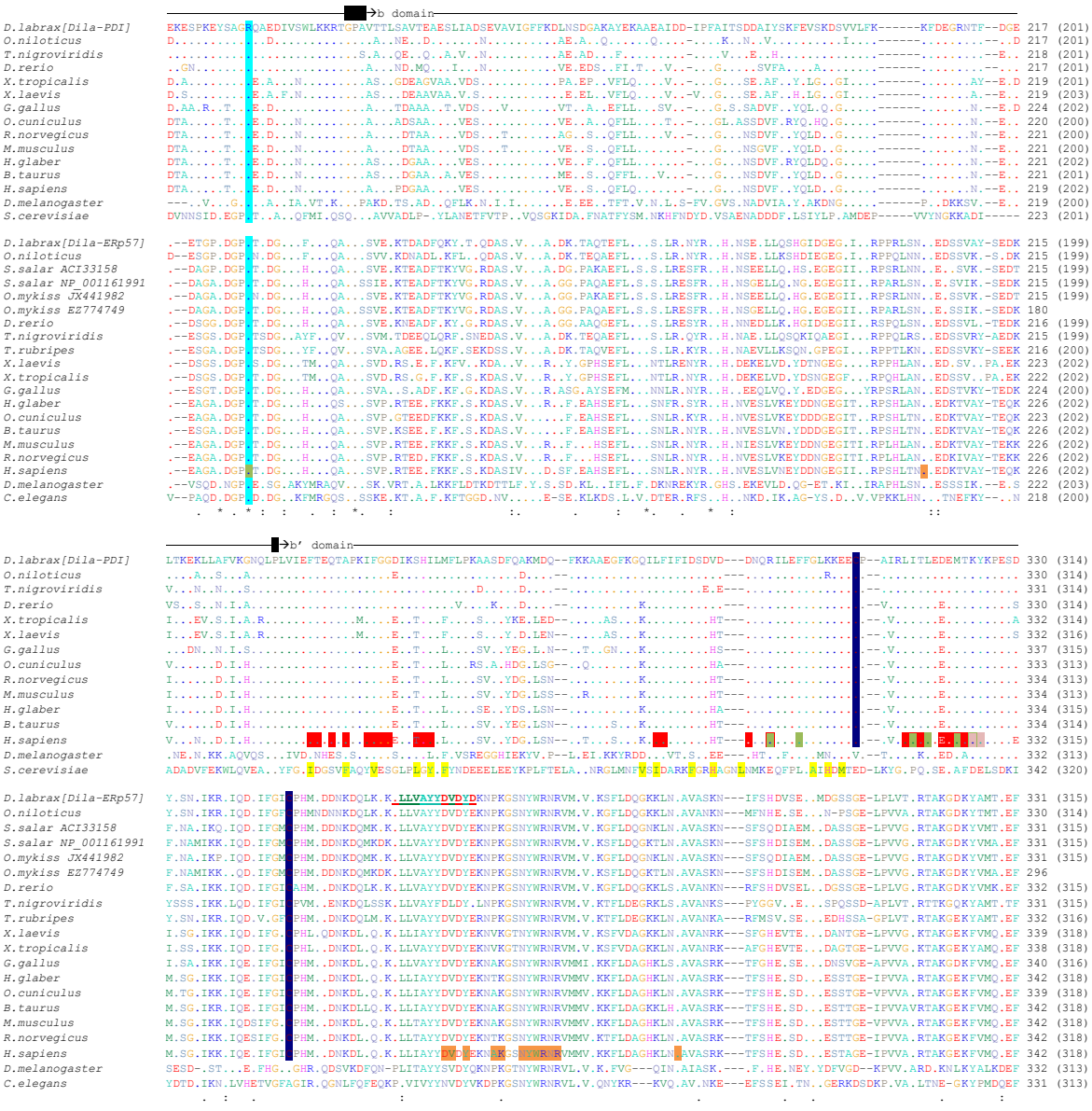


Figure S3 *(continued)*

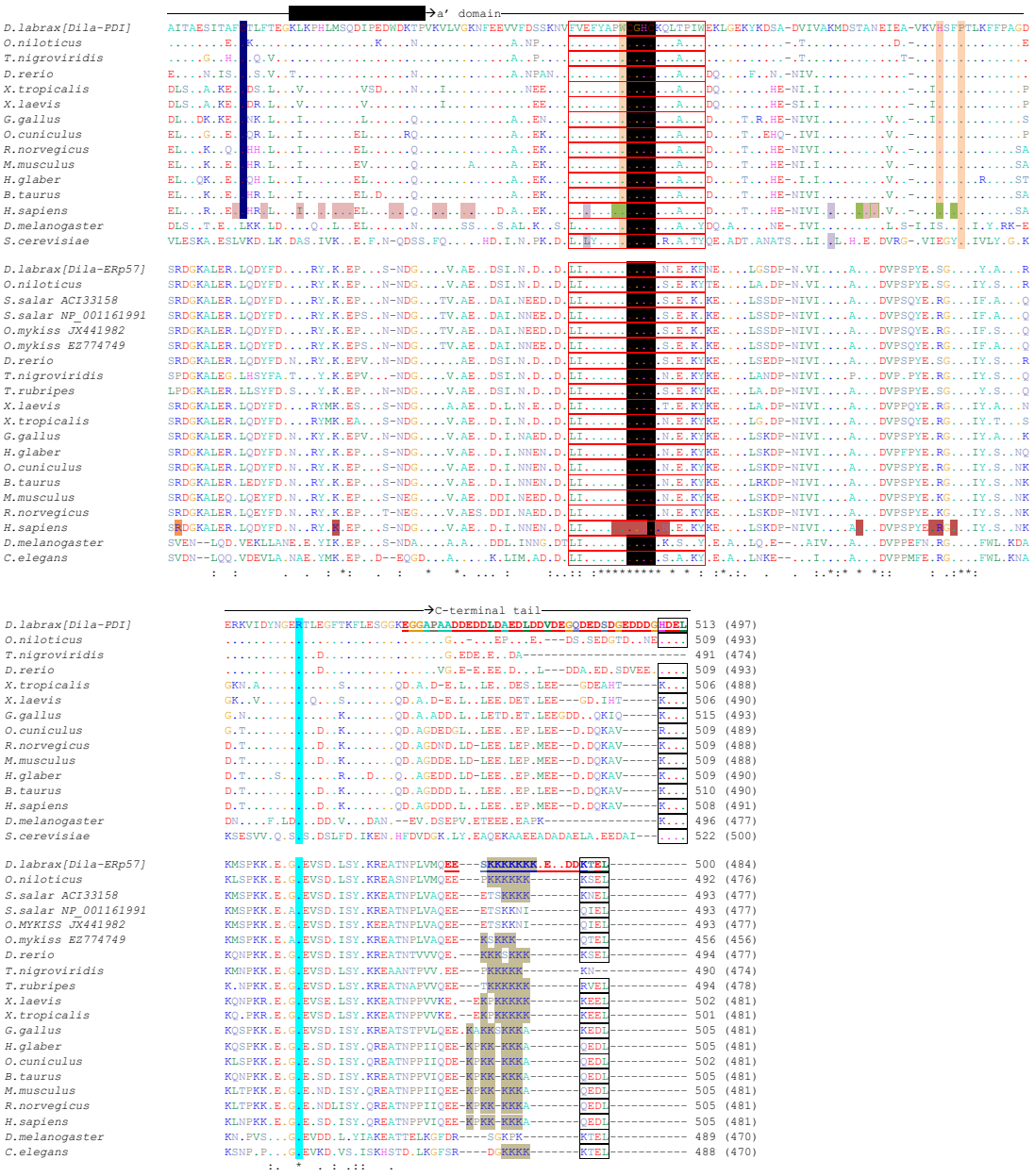


Figure S3 *(continued)*

CHAPTER 7

Concluding remarks and future perspectives

The first reports from fish MHC class I molecules are from the early 1990's [1]. Since then, they have been identified and characterized from a multitude of species, but information on biological functions of these molecules in fish immune system remains limited. The present work describes new and important information about the genetics and the predicted structure of MHC class I molecules in European sea bass (*Dicentrarchus labrax*, L.): heavy chains and β2m, the subunits of the complexes that present peptide antigens to cytotoxic CD8+ T cells, were isolated and characterized; moreover, the sea bass molecules TAP1, TAP2, TPN, CRT, PDI and ERp57, with recognized functions in the mammalian class I antigen presentation pathway, have also been identified and described in detail. Now, the challenge is to use this information to investigate functions of the sea bass immune system, exploring new avenues for vaccine development and to investigate immune responses against specific pathogens.

In the future, it will be important to determine the exact number of MHC-I *loci* present in the sea bass genome and to assign each transcript to a specific *locus*, which would enable an evaluation of the level of polymorphism between alleles. Specific probes for the Cyt-L1 and Cyt-L2 could be used to access the number of genes from each sublineage. Genotyping studies can then be enrolled determining variability of individual *locus* and assessing diversity on populations. Distinct genotyping methods have been developed and used throughout years [2-4], but great advantages are attributed to next generation sequencing (NGS) tools, which widespread use is expected in the future [5, 6]. Such studies will provide a framework for further analyses of MHC infectious disease associations.

Most diseases associated with human MHC (HLA) are autoimmune [7], commonly involving class II genes, but class I associations (e.g. HLA-B27) also exist [8]. In comparison, little is know in relation to resistance or susceptibility to infectious diseases [9]. Convincing evidence for strong reproducible associations between infectious diseases and MHC is coming from studies in non-mammalian species, like those in chickens [10, 11]. Congenic chicken lines varying only in B haplotype have demonstrated the B haplotype effect on resistance to Marek's disease [12]. More recently, significant associations between resistance to infectious diseases caused by viral and bacterial pathogens and MHC class I and class II polymorphism has been reported in Atlanctic salmon [13]. One explanation for such strong correlations is that, despite having multiple *loci*, a single class I *locus* and a single class II *locus* are dominantly expressed [14]. In chicken disease susceptibility has been further linked to co-evolved MHC-I and TAP polymorphisms [15], clearly demonstrating the relevance of genomic linkage of such

genes [16]. Perhaps in species with a high number of class I loci like sea bass, this might be more difficult to demonstrate, but for instance in sticklebacks different haplotypes have been selected by distinct parasites [17].

The MHC-I/peptide recognition by the TCR ultimately regulates the immune responses of CD8+ T cells, which are fundamental during e.g. viral infections. Sea bass is susceptible to viral pathogens, being quite sensitive to viruses of the genus *Betanodavirus* [18-20], the causative agents of viral nervous necrosis (VNN) [21], also known as viral encephalopathy and retinopathy (VER). This infection causes necrosis of the central nervous system (brain, retina and spinal cord), and the clinical signs include lethargy, abnormal behaviour, loss of equilibrium, spiral swimming and darkening of the fish [21]. The disease leads to massive mortality of larvae and juvenile populations in several teleost marine species [22]. Disease manifestation may correlate with innate or adaptive immunity [23-25], but in fact knowledge on the immune responses of this species against the virus is still scarce [26]. The mechanism(s) of infectivity and pathogenesis are unknown, and it remains to be determined whether the virus evades, actively blocks, or uses the molecular recognition factors of the host's immune system [27].

Activation of MHC class I genes during infection with infectious salmon anaemia virus (ISAV) has been reported in Atlantic salmon, suggesting that the virus does not counteract the class I antigen pathway at the transcriptional level [28]. MHC class I activation during an infectious hematopoietic necrosis virus (IHNV) infection in rainbow trout has also been reported [29]. In sea bass, one study presented an integrated view on some innate and adaptive responses of sea bass to VNN infection in naïve fish, identifying cellular and molecular candidate elements whose immunobiology might be modulated by *in vivo* VNN infection [26]. The genes coding for the T cell-specific molecule TCRβ, and for the T cell co-receptors CD4 and CD8α did not show any significant transcriptional variations [26]. Could this be due to some sort of viral strategy targeting the MHC class I pathway genes? Analyzing the behaviour of the key components of the MHC-I pathway may help evaluate the importance of T cell-dependent responses to a given infection, thus opening windows for possible interventions in order to optimise fish defences. The measurement of expressed immune genes, which will provide information on how sea bass respond to particular stimuli/pathogen, can be done on transcripts and at the protein level.

Viruses have evolved strategies to prevent the generation and presentation of antigenic peptides, resulting in their ability to escape from the immune system. Interestingly, viruses can target many crucial steps of antigen processing simultaneously, and many of the viral proteins that target antigen presentation directly assault the peptide-loading complex *per se* [30]. Preventing this binding will result in a reduced pool of

peptides in the ER available for association with MHC-I, ultimately resulting in a reduced level of MHC-I cell surface expression [30]. E.g. there are viral proteins that target TAP, preventing it to bind peptides and ultimately blocking the translocation of peptides into the ER [31, 32]. Another way to prevent peptide translocation is to prevent ATP binding to the TAP subunits [33]. A combination of both mechanisms has also been described [34]. Abrogation of the MHC/TPN interaction was reported, preventing MHC integration in the PLC [35]. The quality control mediated by tapasin in the PLC can also be negatively affected [36]. Finally, the different elements of the PLC can be targeted to degradation by interaction with viral proteins [37-39].

The current work has provided basic knowledge for the production of antibodies, namely monospecific polyclonal antibodies (m-pAbs) against specific peptides from the identified molecules. Such tools have been generated for sea bass heavy chain (α3 region), β2m, TPN, TAP1 and TAP2, and are currently being tested (preliminary data detailed in Appendix). A similar strategy should be employed in order to produce m-pAbs for Dila-CRT, -PDI and -ERp57. Nevertheless, it would be worth testing commercial antibodies against CRT, ERp57 and PDI from mammalian species, given the high level of identity observed between them (~ 70% with human).

Collectively, our data point to an evolutionary conservation of basic structures and functions, with predicted orthology of the sea bass molecules to their mammalian equivalents; in the particular case of the class I α chains, orthology is predicted to Actinopterygii and salmonid counterparts only, which is not surprising, since orthology is not broadly observed even among mammalian MHC class I genes. Nonetheless, further specificities were observed, such as the presence of two types of HC transcripts with divergent cytoplasmic and 3' UTRs, a 13-exon TAP1 gene, and a TAP2 isoform with incorporated intron 2, whose relevance remains to be determined. The highest levels of conservation were observed among the common chaperones CRT, ERp57 and PDI, which are known to be involved in a variety of other functions beyond the class I pathway, and are probably not subjected to the same selective pressures of the MHC-encoded proteins. The comparative structural analyses that was done in the previous chapters for each molecule individually, suggests that the described multiple interactions between them in the context of the PLC (in mammals) may be overall evolutionarily conserved. Significant conservation of residues in the interplay was observed for all except the TPN:ERp57 interaction. Dila-TPN, as other non-mammalian TPNs, lacks the cysteine that should be involved in the covalent association with ERp57 and many of the residues that further support this association are also not conserved. Hence, important questions are raised, such as: are these key players still capable of forming a stable conjugate in sea

bass and other non-mammalian vertebrates? How do they associate? Antibodies, through co-immunoprecipitation experiments could help answering these questions, as well as confirm (or not) all the other predicted interactions among the PLC members. Additionally, the experimental resolution of the structures of these molecules, either isolated or in complexes, would also bring valuable insights.

Since the first crystallographic study revealed a short peptide bound within the groove of the molecule [40], the structure of peptide-binding specificity of MHC class I has been analyzed extensively in human and mouse. Distinct MHC HCs molecules have distinct peptide binding specificities in such a way that peptide-binding motifs could be defined. Among non-mammalian vertebrates, only in chicken there are experimental crystallographic data of the full MHC class I complex [41] and sequences of individual peptides and peptide pools leading to peptide motifs for the dominantly expressed classical MHC class I molecules (except for the crystallized molecule) [42, 43]. In fish there are no crystallographic models of MHC complexes, and only one report with data on peptide-binding specificity: class I peptide-MHC ligand-binding motif has been reported in Atlantic salmon [44]. Recently, grass carp classical class I *locus* and β2m were expressed and refolded *in vitro* (viral immunogenic peptides predicted by bioinformatics), and tetramer techniques were used to identify the CTL response [45]. These studies might contribute to the understanding of viral defence mechanisms in general and to the design of peptide vaccines against viral diseases in particular [46].

The current work on sea bass MHC class I molecules has enabled to start *in vitro* refolding experiments aiming at crystallizing a sea bass MHC class I trimeric complex (Appendix). We attempted to refold several recombinant proteins of Dila-UA molecules with Dila-β2m or human β2m in the presence of different synthetic peptides. In none of the experiments did we get a trimeric complex. Other peptides capable of triggering T cell activation should be investigated. The nodavirus capsid or coat protein is the main viral antigen employed shown to be immunogenic in fish [47-50]. Accordingly, the produced recombinant nodavirus capsid protein, which is capable of self-assembly to form virus-like particles, has been shown to be immunogenic and induce a protective immune response in sea bass [51]. With these premises, studies on putative peptide ligands may have a value for both the *in vitro* refolding experiments and for understanding peptide binding to Dila-UA molecules.

Alternatively, to identify peptide binders for DILA-UA sequences, we have transfected a human lymphoblastoid cell line with some of the sea bass representatives of both Cyt-L1 and Cyt-L2 (detailed information in Appendix). Full Dila-UA transcripts have been used in combination with GFP. The transfections worked, but none of the molecules

reached the surface. Possibly the fish molecules did not pass the quality machinery of the mammalian host cells.

This strategy would also enable to clarify putative distinct functions of Cyt-L1 and Cyt-L2 molecules. The distinct specificities observed within the cytoplasmic regions of the characterized class I α chains, including presence of Y residues and LI motifs with roles in endosomal targeting, are highly suggestive of putative different functions. It is mandatory to clarify this in near future. To overcome this perhaps working with chimeras of sea bass and human HCs would make it possible to track the two types of sea bass cytoplasmic tails to distinct organelles/pathways.

In conclusion, there is still a lot to investigate about antigen presentation in fish. Indeed, many questions are still unanswered [52]. The limitations include the lack of specific antibodies available, which is not completely understood. It has been proposed that the heavily glycosylated surface of fish cells results in production of antibodies that bind to the glycosylated surface of all fish cells, rather than the specific antigen target [52]. Importantly, comparative genomic analyses are giving insights on both similarities and species-specificities of immune systems in lower vertebrates [53]. This type of approach (genomic immunology) has been impacting the understanding of mammalian immune systems, providing new ways to overcome defective or inadequate immune functions [53]. Hopefully, advances in this field will help design antiviral strategies for the farming of European sea bass, contributing to the development of vaccines, selection of disease-resistant breeds and broadly to a better understanding of the evolution of the immune system [54].

References

1. Hashimoto K, Nakanishi T, Kurosawa Y. Isolation of carp genes encoding major histocompatibility complex antigens. Proc Natl Acad Sci U S A. 1990 87:6863-7.
2. Babik W. Methods for MHC genotyping in non-model vertebrates. Mol Ecol Resour. 2010 10:237-51.
3. Lenz TL, Eizaguirre C, Becker S, Reusch TB. RSCA genotyping of MHC for high-throughput evolutionary studies in the model organism three-spined stickleback Gasterosteus aculeatus. BMC Evol Biol. 2009 9:57.
4. Dunn PP. Human leucocyte antigen typing: techniques and technology, a critical appraisal. Int J Immunogenet. 2011 38:463-73.
5. Erlich H. HLA DNA typing: past, present, and future. Tissue Antigens. 2012 80:1-11.
6. De Santis D, Dinauer D, Duke J, Erlich HA, Holcomb CL, Lind C, et al. 16(th) IHIW : review of HLA typing by NGS. Int J Immunogenet. 2013 40:72-6.
7. Guarene M, Capittini C, De Silvestri A, Pasi A, Badulli C, Sbarsi I, et al. Targeting the Immunogenetic Diseases with the Appropriate HLA Molecular Typing: Critical Appraisal on 2666 Patients Typed in One Single Centre. Biomed Res Int. 2013 2013:904247.
8. Martinez-Borra J, Gonzalez S, Lopez-Vazquez A, Gelaz MA, Armas JB, Kanga U, et al. HLA-B27 alone rather than B27-related class I haplotypes contributes to ankylosing spondylitis susceptibility. Hum Immunol. 2000 61:131-9.
9. Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. Clin Microbiol Rev. 2009 22:370-85, Table of Contents.
10. Bacon LD, Hunt HD, Cheng HH. A review of the development of chicken lines to resolve genes determining resistance to diseases. Poult Sci. 2000 79:1082-93.
11. Lamont SJ. The chicken major histocompatibility complex and disease. Rev Sci Tech. 1998 17:128-42.
12. Bacon LD. Influence of the major histocompatibility complex on disease resistance and productivity. Poult Sci. 1987 66:802-11.
13. Grimholt U, Larsen S, Nordmo R, Midtlyng P, Kjoeglum S, Storset A, et al. MHC polymorphism and disease resistance in Atlantic salmon (Salmo salar); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics. 2003 55:210-9.
14. Kaufman J. Antigen processing and presentation: Evolution from a bird's eye view. Mol Immunol. 2013 55:159-61.

15. Zhang J, Chen Y, Qi J, Gao F, Liu Y, Liu J, et al. Narrow groove and restricted anchors of MHC class I molecule BF2*0401 plus peptide transporter restriction can explain disease susceptibility of B4 chickens. J Immunol. 2012 189:4478-87.
16. Walker BA, Hunt LG, Sowa AK, Skjodt K, Gobel TW, Lehner PJ, et al. The dominantly expressed class I molecule of the chicken MHC is explained by coevolution with the polymorphic peptide transporter (TAP) genes. Proc Natl Acad Sci U S A. 2011 108:8396-401.
17. Eizaguirre C, Lenz TL, Kalbe M, Milinski M. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat Commun. 2012 3:621.
18. Frerichs GN, Rodger HD, Peric Z. Cell culture isolation of piscine neuropathy nodavirus from juvenile sea bass, Dicentrarchus labrax. J Gen Virol. 1996 77 (Pt 9):2067-71.
19. Peducasse S, Castric J, Thiery R, Jeffroy J, Le Ven A, Baudin Laurencin F. Comparative study of viral encephalopathy and retinopathy in juvenile sea bass Dicentrarchus labrax infected in different ways. Dis Aquat Organ. 1999 36:11-20.
20. Breuil G, Mouchel O, Fauvel C, Pepin JF. Sea bass Dicentrarchus labrax nervous necrosis virus isolates with distinct pathogenicity to sea bass larvae. Dis Aquat Organ. 2001 45:25-31.
21. Bovo G, Nishizawa T, Maltese C, Borghesan F, Mutinelli F, Montesi F, et al. Viral encephalopathy and retinopathy of farmed marine fish species in Italy. Virus Res. 1999 63:143-6.
22. Munday BL, Kwang, J., Moody, N. Betanodavirus infections of teleost fish: a review. . J Fish Dis. 2002 25:127-42.
23. Lu MW, Chao YM, Guo TC, Santi N, Evensen O, Kasani SK, et al. The interferon response is involved in nervous necrosis virus acute and persistent infection in zebrafish infection model. Mol Immunol. 2008 45:1146-52.
24. Poisa-Beiro L, Dios S, Montes A, Aranguren R, Figueras A, Novoa B. Nodavirus increases the expression of Mx and inflammatory cytokines in fish brain. Mol Immunol. 2008 45:218-25.
25. Novel P, Fernandez-Trujillo MA, Gallardo-Galvez JB, Cano I, Manchado M, Buonocore F, et al. Two Mx genes identified in European sea bass (Dicentrarchus labrax) respond differently to VNNV infection. Vet Immunol Immunopathol. 2013.
26. Scapigliati G, Buonocore F, Randelli E, Casani D, Meloni S, Zarletti G, et al. Cellular and molecular immune responses of the sea bass (Dicentrarchus labrax) experimentally infected with betanodavirus. Fish Shellfish Immunol. 2010 28:303-11.

27. Poisa-Beiro L, Dios S, Ahmed H, Vasta GR, Martinez-Lopez A, Estepa A, et al. Nodavirus infection of sea bass (Dicentrarchus labrax) induces up-regulation of galectin-1 expression with potential anti-inflammatory activity. J Immunol. 2009 183:6600-11.
28. Jorgensen SM, Hetland DL, Press CM, Grimholt U, Gjoen T. Effect of early infectious salmon anaemia virus (ISAV) infection on expression of MHC pathway genes and type I and II interferon in Atlantic salmon (Salmo salar L.) tissues. Fish Shellfish Immunol. 2007 23:576-88.
29. Hansen JD, La Patra S. Induction of the rainbow trout MHC class I pathway during acute IHNV infection. Immunogenetics. 2002 54:654-61.
30. Roder G, Geironson L, Bressendorff I, Paulsson K. Viral proteins interfering with antigen presentation target the major histocompatibility complex class I peptide-loading complex. J Virol. 2008 82:8246-52.
31. York IA, Roop C, Andrews DW, Riddell SR, Graham FL, Johnson DC. A cytosolic herpes simplex virus protein inhibits antigen presentation to CD8+ T lymphocytes. Cell. 1994 77:525-35.
32. Fruh K, Ahn K, Djaballah H, Sempe P, van Endert PM, Tampe R, et al. A viral inhibitor of peptide transporters for antigen presentation. Nature. 1995 375:415-8.
33. Kyritsis C, Gorbulev S, Hutschenreiter S, Pawlitschko K, Abele R, Tampe R. Molecular mechanism and structural aspects of transporter associated with antigen processing inhibition by the cytomegalovirus protein US6. J Biol Chem. 2001 276:48031-9.
34. Hislop AD, Ressing ME, van Leeuwen D, Pudney VA, Horst D, Koppers-Lalic D, et al. A CD8+ T cell immune evasion protein specific to Epstein-Barr virus and its close relatives in Old World primates. J Exp Med. 2007 204:1863-73.
35. Bennett EM, Bennink JR, Yewdell JW, Brodsky FM. Cutting edge: adenovirus E19 has two mechanisms for affecting class I MHC expression. J Immunol. 1999 162:5049-52.
36. Park B, Kim Y, Shin J, Lee S, Cho K, Fruh K, et al. Human cytomegalovirus inhibits tapasin-dependent peptide loading and optimization of the MHC class I peptide cargo for immune evasion. Immunity. 2004 20:71-85.
37. Boname JM, de Lima BD, Lehner PJ, Stevenson PG. Viral degradation of the MHC class I peptide loading complex. Immunity. 2004 20:305-17.
38. Lybarger L, Wang X, Harris MR, Virgin HWt, Hansen TH. Virus subversion of the MHC class I peptide-loading complex. Immunity. 2003 18:121-30.

39. Park B, Lee S, Kim E, Cho K, Riddell SR, Cho S, et al. Redox regulation facilitates optimal peptide selection by MHC class I during antigen processing. Cell. 2006 127:369-82.
40. P.J. Bjorkmann MAS, B. Samraoui, W.S. Bennett, J.L. Strominger & D.C. Wiley. Structure of the human class I histocompatibility complex, HLA-A2. Nature. 1987 329:506-12.
41. Koch M, Camp S, Collen T, Avila D, Salomonsen J, Wallny HJ, et al. Structures of an MHC class I molecule from B21 chickens illustrate promiscuous peptide binding. Immunity. 2007 27:885-99.
42. Kaufman J, Volk H, Wallny HJ. A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. Immunol Rev. 1995 143:63-88.
43. Wallny HJ, Avila D, Hunt LG, Powell TJ, Riegert P, Salomonsen J, et al. Peptide motifs of the single dominantly expressed class I molecule explain the striking MHC-determined response to Rous sarcoma virus in chickens. Proc Natl Acad Sci U S A. 2006 103:1434-9.
44. Zhao H, Hermsen T, Stet RJ, Skjodt K, Savelkoul HF. Peptide-binding motif prediction by using phage display library for SasaUBA*0301, a resistance haplotype of MHC class I molecule from Atlantic Salmon (Salmo salar). Mol Immunol. 2008 45:1658-64.
45. Chen W, Jia Z, Zhang T, Zhang N, Lin C, Gao F, et al. MHC class I presentation and regulation by IFN in bony fish determined by molecular analysis of the class I locus in grass carp. J Immunol. 2010 185:2209-21.
46. Chistiakov DA, Hellemans B, Volckaert FA. Review on the immunology of European sea bass Dicentrarchus labrax. Vet Immunol Immunopathol. 2007 117:1-16.
47. Emmenegger E, Huang C, Landolt M, LaPatra S, Winton JR. Immune response to synthetic peptides representing antigenic sites on the glycoprotein of infectious hematopoietic necrosis virus. Vet Res. 1995 26:374-8.
48. Cain KD, LaPatra SE, Shewmaker B, Jones J, Byrne KM, Ristow SS. Immunogenicity of a recombinant infectious hematopoietic necrosis virus glycoprotein produced in insect cells. Dis Aquat Organ. 1999 36:67-72.
49. Yuasa K KI, Roza D, Mori K, Katata M, Nakai T. Immune response of humpback grouper, Cromileptes altivelis (Valenciennes) injected with the recombinant coat protein of betanodavirus. Journal of Fish Diseases. 2002 25:53-6.
50. Coeurdacier JL, Laporte F, Pepin JF. Preliminary approach to find synthetic peptides from nodavirus capsid potentially protective against sea bass viral encephalopathy and retinopathy. Fish Shellfish Immunol. 2003 14:435-47.

51. Thiery R, Cozien J, Cabon J, Lamour F, Baud M, Schneemann A. Induction of a protective immune response against viral nervous necrosis in the European sea bass Dicentrarchus labrax by using betanodavirus virus-like particles. J Virol. 2006 80:10201-7.
52. Bassity E, Clark TG. Functional identification of dendritic cells in the teleost model, rainbow trout (Oncorhynchus mykiss). PLoS One. 2012 7:e33196.
53. Boehm T, Iwanami N, Hess I. Evolution of the immune system in the lower vertebrates. Annu Rev Genomics Hum Genet. 2012 13:127-49.
54. Zhu LY, Nie L, Zhu G, Xiang LX, Shao JZ. Advances in research of fish immune-relevant genes: a comparative overview of innate and adaptive immunity in teleosts. Dev Comp Immunol. 2013 39:39-62.

Appendix

Preliminary studies to further characterize sea bass (*Dicentrarchus labrax*, L.) MHC class I complexes and components of the processing pathway

I. Production of monospecific polyclonal antibodies against sea bass (*Dicentrarchus labrax*, L.) MHC class I heavy chain, β2m, tapasin, TAP1 and TAP2

Based on the work presented in this thesis, monospecific polyclonal antibodies could be produced against some of the molecules composing the peptide loading complex. Due to the limited time available, within the time-frame of this thesis, only very preliminary tests could be performed, which are briefly summarised below.

Materials and Methods

Peptides

Short peptides from the sea bass molecules MHC-I HC, β2m, TPN, TAP1 and TAP2 were used as immunogens in rabbits (Table 1). Selection of peptides was based on antigenicity, accessibility, solubility and epitope prediction. Peptide synthesis, immunization of animals and affinity purification was done at Davids Biotechnologie GmbH.

Lymphocyte isolation

Sea bass lymphocytes were isolated either from blood or from distinct organs. Blood was collected into a heparinised syringe and diluted 1:1 in sea bass PBS (10 mM $NaH_2PO_4.H_2O$ pH 7.2, 184 mM NaCl). After centrifuging 15 min at 300 g and 4 ºC, the supernatant was collected into another tube and centrifuged for (10 min, 550 g, 4 ºC). The pellet was resuspended in 3 mL of sbPBS. This cell suspension was added on top of an equal volume of Lymphoprep and centrifuged (30 min, 1100 g 4 ºC, no brake). The layer containing lymphocytes was carefully collected with a Pasteur pipette and washed in sbPBS, by centrifuging (10 min, 550 g, 4 ºC). After resuspending in 2 mL sbPBS and counting, cells were adjusted to 1 x 10^6, pelletized by centrifuging (10 min, 550 g, 4 ºC), and stored at -20 ºC.

Organs (thymus, spleen, head kidney, gills, or intestine) were homogenized by passing them through a cell filter and resuspended in sbPBS. Cell suspensions were washed twice by centrifugation (10 min, 550 g, 4 ºC). Pellets were resuspended in sbPBS and added on top of an equal volume of Lymphoprep. Lymphocytes were collected, washed twice and resuspended as described above. Cells were counted and pellets

stored.

SDS-PAGE

Samples were prepared by addition of gel loading buffer (GLB) and a reducing agent (β-mercaptoethanol), adjusted with distilled water to a final volume of 30 μL, boiled for 5 min, and run in a SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis). Depending on the molecules, distinct percentages of resolving gels were used.

Western Blot

Samples in gels were transferred to nitrocellulose membranes (transfer buffer: 25 mM Tris, 192 mM glycine, 20% methanol), for 1 hour at 19V. Membranes were Ponceau stained, scanned and blocked by incubating for 1 hour at RT, or overnight at 4 ºC, in 5% low fat milk in T-TBS (20 mM Tris Base, 137 mM NaCl, 0.1% v/v Tween 20). Incubation with variable concentrations of primary antibody (IgG purified serum) was performed in blocking buffer overnight at 4 ºC. Washes (5 x 5 min) were made in T-TBS at RT with shaking. Incubation with secondary antibody (sheep anti-rabbit HRP-conjugated 1:10000) was done four 1 hour at RT and washes performed as before. For detection, ECL (Supersignal WestDura Extended Duration or Supersignal WestPico Chemiluminescent Substrate, Thermo Scientific) was used and the films exposed for variable times.

Results

Dila-UA antiserum

The anti-serum was initially tested on the extracellular portion (V1-K276, mature protein; Fig. S3 Chapter 2) of His-tagged Dila-UA*12(C91) produced as recombinant protein and purified by nickel affinity chromatography. Different dilutions of the antiserum were used (1:1000, 1:5000, 1:10000), yielding no band at all in the negative control (without primary antibody), and a well-defined band at the expected size (~33 kDa) with decreasing intensities towards the highest dilution (not shown). The recombinant was used as a positive control (+) in further tests with fish cell lysates from different organs. Again different dilutions of the primary antibody were tested. In Figure 1, are depicted the WB results in which a 1:2500 dilution of the Dila-UA antiserum was used in fish samples from different organs. Three bands can be clearly seen close to the 45 kDa Mw marker on the thymus sample from one fish (expected size 38-43 kDa, Table 1). Exposing the film for longer periods made all the three bands visible in the sample of thymus from the other individual fish as well (not shown). No bands could be seen in samples from other organs, except a low Mw one (< 14.4 kDa), not expected, in a head kidney (HK) sample.

Taking into account the variable size of the MHC-I HCs (Chapter 2, Table 1) and that the thymus is organ where highest expression was observed (Chapter 2, Fig. 8), this antibody seems promising as a sea bass MHC-I marker. However more tests need to be done in order to optimise it.

Dila-β2m antisera

Two antisera (#1 and #2; Table 1) were produced for Dila-β2m. Both anti-sera were initially tested on bacterial lysates containing the recombinant denatured histidine-purified Dila-β2m (I6V-M101, mature protein; Fig. 4 Chapter 2) using different dilutions (1:1000, 1:5000, 1:10000). Antiserum #1 yielded no band at all in the negative control (no primary), and a well-defined band at the expected size with decreasing intensities towards the highest dilution (Fig. 2, upper panel). No bands could be observed when antiserum #2 was used (Fig. 2, upper panel). Recombinant Dila-β2m was used as a positive control (+) in further tests with fish cell lysates from different organs, using only antiserum #1. Again different dilutions of the primary antibody were tested. Although some bands can be seen close to the size of the recombinant protein in the 1:1000 dilution (Fig. 2, lower panel), with emphasis for a stronger one in the gills sample, further tests are needed to conclude whether these bands represent a specific reaction and the conditions improved such that the antibody can be used in functional studies.

Dila-TPN antiserum

For Dila-TPN there is no recombinant protein available, so the tests were performed immediately on sea bass cell lysates. Tests were initially performed using different dilutions of the primary antibody. Figure 3 displays the results when a 1:10000 dilution was tested in samples from three individuals. The expected size is 45 kDa, and some bands can be observed close to this size, the strongest one in a gill sample (Fig. 3). The weaker bands on the film may correspond to the stronger ones seen in the Ponceau-stained membrane (Fig. 3), in which case they might not be specific. Further tests are necessary to clarify this.

Dila-TAP1 antisera

Three antisera (#0, #4 and #5; Table 1) were produced for Dila-TAP1. As no recombinant TAP1 protein was available, tests were performed on sea bass cell lysates from different organs, using different dilutions of the primary antibody. Antiserum #0 always stained many bands, independently of the dilution used, and no band of 83 kDa (expected size) could be observed (Fig. 4A). Antiserum #4 stained fewer bands, but still none close to the expected size (Fig. 4B). Antiserum #5 was the only staining a band close to the expected size (83 kDa), although lower and higher Mw bands with low intensity are also seen (Fig. 4C). Thus, antiserum #5 appears as the most promising antibody against Dila-TAP1, which conditions may well be worth of further optimisation.

Dila-TAP2 antisera

Three antisera (#0, #1 and #4; Table 1) were produced for Dila-TAP2. As no recombinant TAP2 protein was available, tests were performed on sea bass cell lysates from different organs, using different dilutions of the primary antibody. Antiserum #0 only stained bands <45 kDa, when the expected size is 80 kDa (Fig. 5A). Antiserum #1 did not work at all, independently of the dilution used (Fig 5B). Antiserum #4 stained a single band (1:1000 dilution) close to the expected size (80 kDa) in spleen samples (Fig. 5B). A 1:2500 dilution has also been shown to work (not shown). Hence, antiserum #4 appears as a very promising antibody to TAP2. Tests using samples from different organs will help defining the potential use of this antibody in functional tests.

Table 1

Features of the sera produced against the sea bass molecules.

Sea bass molecule	Type	Antigen	Region	Expected Mw	Recombinant protein
Dila-UA	Rabbit IgG	VSLLQKTPSSPVSCHATGFY	IgSf-C (α3 region)	38.3-43.9 kDa	Yes (33 kDa)
Dila-β2m	Rabbit IgG	#1 SDNPKITIKPKVQVYSRDPGE #2 DGEEIQNAKQTDLAF	N-term mature IgSf-C	11.6 kDa	Yes (12.8 kDa)
Dila-TPN	Rabbit IgG	PLSLELSWEFKGSDGKSR	IgSf-C	45 kDa	No
Dila-TAP1	Rabbit IgG	#0 YAELVRKQNMGFHRKEEEAK #4 EDKLVLKDLSLEI #5 EKANHIVVLSDGTVKEEGSH	C-terminus NBD C-terminus	83 kDa	No
Dila-TAP2	Rabbit IgG	#0 KGSYYKLREKLFTEGNSPQR #1 TIPDTDDAANGKEKKQKA #4 EKADQIVVVSGGRVQERGTHQE	C-terminus N domain of TMD C-terminus	80 kDa	No

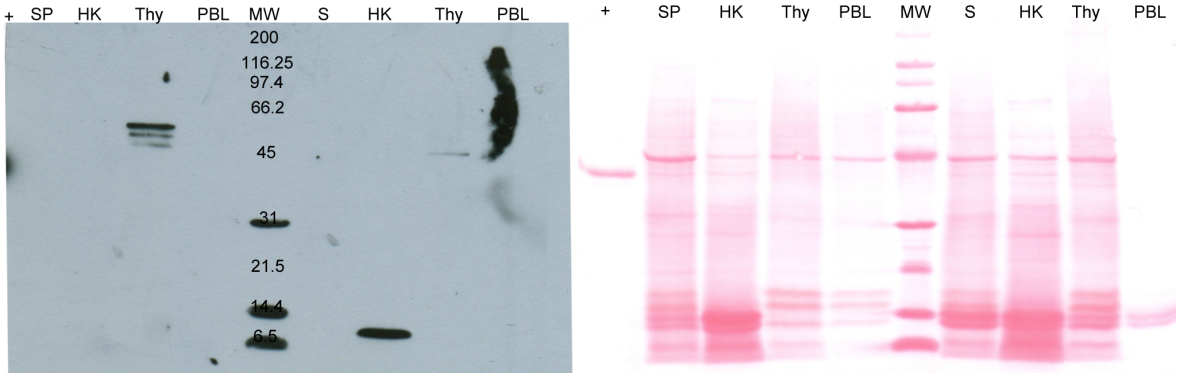


Figure 1. Western blot using anti-Dila-UA serum. Antiserum tested in 1:2500 dilution on sea bass cell lysates from two different fish (left and right to the MW, respectively). Sensitive ECL plus 1 hour film exposure time was used. Film is shown at left and Ponceau stained membrane at right. Sizes of MW are indicated. +, recombinant Dila-UA*12; SP, spleen; HK, head kidney; Thy, thymus; PBL, peripheral blood lymphocytes.

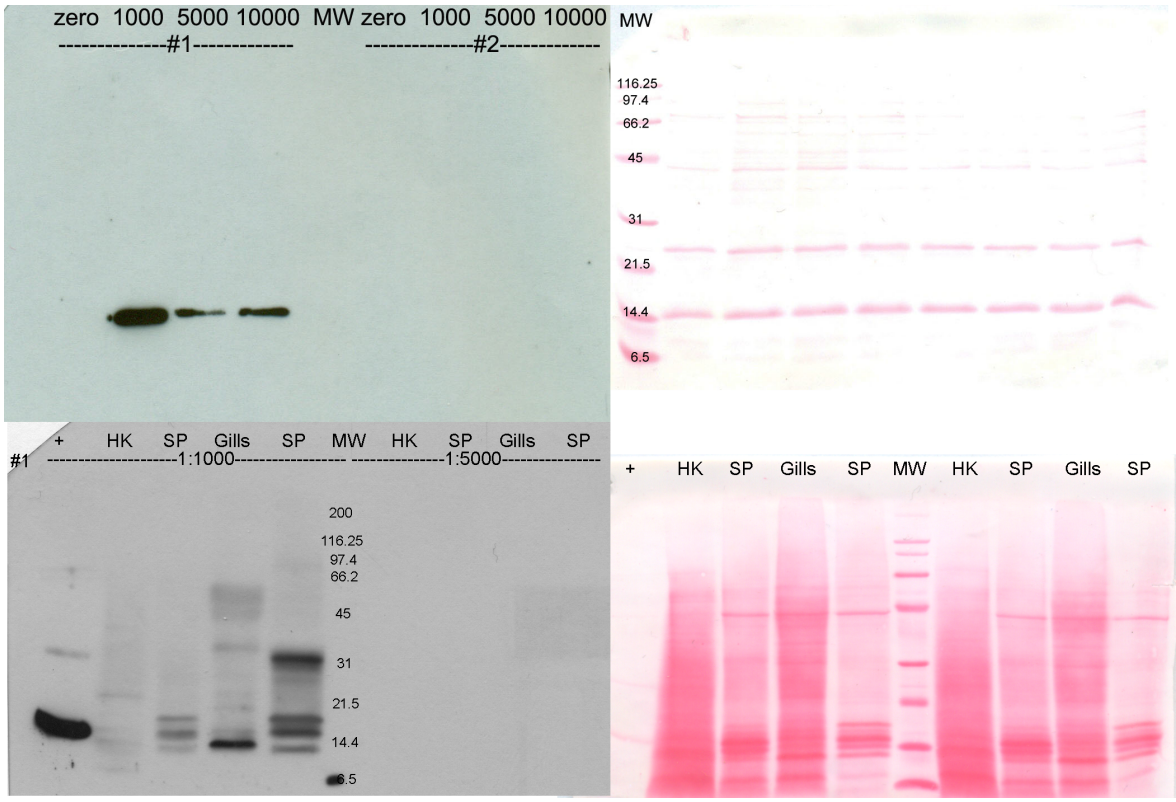


Figure 2. Western blot using anti-β2m sera. The two antisera (#1 and #2) were tested in different dilutions (1:1000, 1:5000, 1:10000) on bacterial lysates containing recombinant β2m (upper panels) and on sea bass cell lysates (lower panels). Normal ECL plus 5 min exposure time and sensitive ECL plus 10 sec exposure time were used for the recombinant and the fish cells, respectively. Films are shown at left and Ponceau stained membranes on right. Sizes of MW are indicated. Zero, no primary antibody; +, bacterial cell lysates with recombinant β2m; HK, head kidney; SP, spleen.

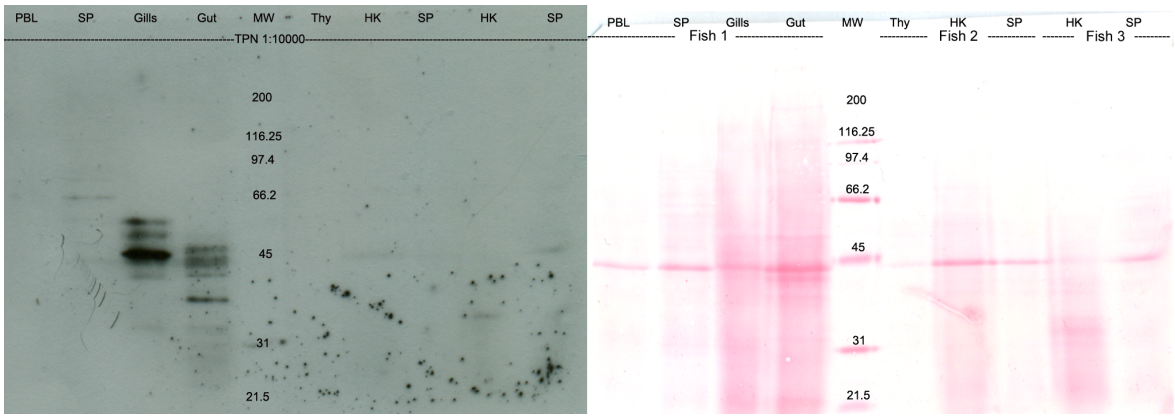


Figure 3. Western blot using anti-TPN serum. Antiserum tested in 1:10000 dilution on sea bass cell lysates from three different fish (1, 2, 3) as detailed in the membrane. The expected size is 45 kDa. Sensitive ECL plus 30 min film exposure time were used. Film is shown at left and Ponceau stained membrane at right. Sizes of MW are indicated. PBL, peripheral blood lymphocytes; HK, head kidney; SP, spleen.

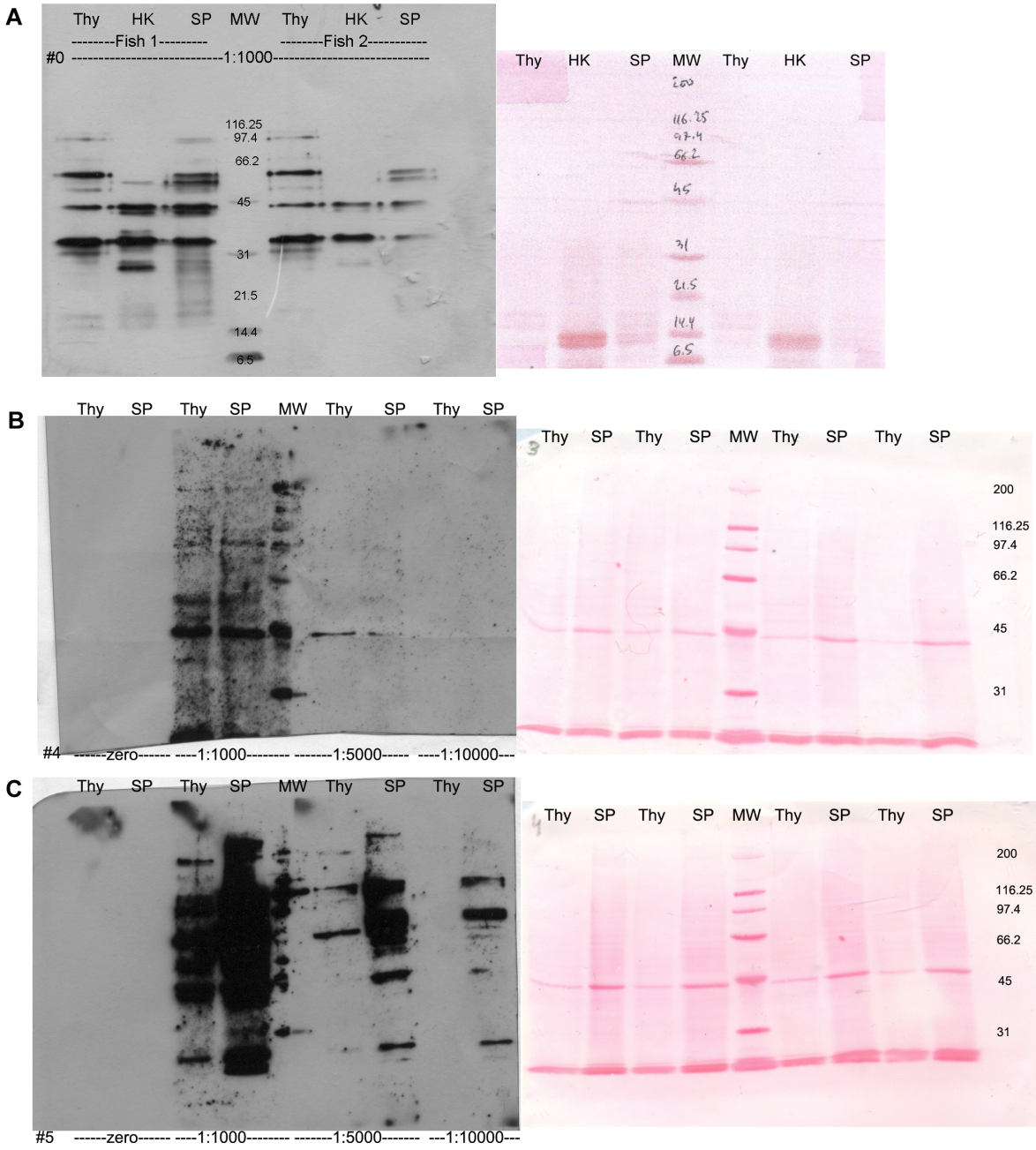


Figure 4. Western blot using anti-TAP1 sera. The three antisera (TAP1#0, TAP1#4 and TAP1#5) were tested on sea bass cell lysates. from different organs. Expected size is 83 kDa. **(A)** TAP1#0 tested with a 1:1000 dilution is shown. **(B)** TAP1#4 **(C)** TAP1#5, both tested in different dilutions (1:1000, 1:5000, 1:10000). Sensitive ECL and 10 sec (A) or 1 min (B and C) exposure times on film were used. Films are shown at left and Ponceau stained membranes on right. Sizes of MW are indicated. Zero, no primary antibody; Thy, thymus; SP, spleen.

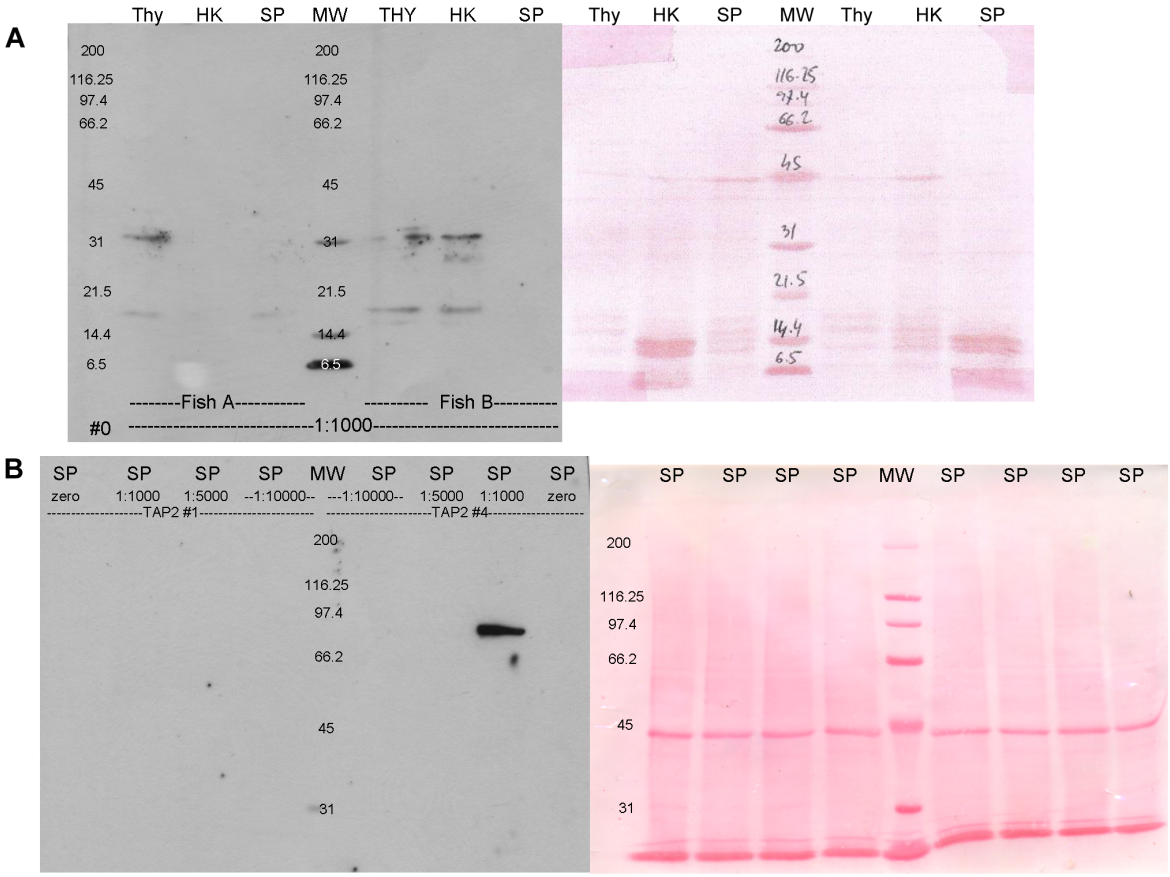


Figure 5. Western blot of anti-TAP2 sera. The three antisera (TAP2#0, TAP2#1 and TAP2#4) were tested on sea bass cell lysates from different organs. The expected size is 80 kDa. **(A)** TAP2#0 tested with a 1:1000 dilution is shown. **(B)** TAP2#1 (left side of MW) and TAP2#4 (right side of MW), were both tested in different dilutions (1:1000, 1:5000, 1:10000). Sensitive ECL and 1 min (A) or 5 min (B) exposure times on film were used. Films are shown at left and Ponceau stained membranes on right. Sizes of MW are indicated. Zero, no primary antibody; Thy, thymus; HK, head kidney; SP, spleen.

II. Preliminary studies aiming at determination of the three-dimensional structure of sea bass (*D. labrax*, L.) MHC class I complexes

A significant effort has been devoted on attempting to determine the three-dimensional structure of sea bass MHC class I complexes. Although the 3D structure has not been determined within the time-frame of the working plan, several techniques have been applied and the results obtained worth to be considered as a starting point for defining conditions to grow suitable crystals that can be used for disclosing the 3D structure of the sea bass MHC class I complexes. This work is summarized below.

Materials and Methods

Recombinant protein production: from vector design to protein storage

The extracellular portion of three sea bass heavy chain (HC) transcripts (Dila-UA*06, *09 or *12; see chapter 2), encoding residues V1-K276 (mature protein; Fig. S3 Chapter 2), after PCR incorporation of a methionine at the 5' end, were cloned in pET28a vector (using the NcoI/XhoI restriction sites) for recombinant expression in *E. coli* BL21. Sea bass β2-microglobulin (β2m) encoding residues I6-M101 (mature protein; Fig. 4 Chapter 2) – starting in the equivalent position to the human recombinant protein – was similarly expressed in BL21 after PCR incorporation of a methionine at the 5' end and substitution of isoleucine-6 for valine (necessary for the correct NcoI restriction site), and ligation to the mentioned pET28a vector (also using the NcoI/XhoI restriction sites). All constructs were verified by sequencing.

Inductions were made with 0.5 mM IPTG at 37 ºC. As previously described for similar molecules of other organisms, recombinant expression resulted in accumulation of insoluble material in inclusion bodies. Inclusion bodies were isolated by cell lysis and repeated washes, essentially as described by Nagai and Thorgensen [1], and solubilised under denaturing conditions (8 M urea buffer). The recombinant proteins were purified through the use of the C-terminal 6-histidine tag, using HIS Select Nickel Affinity Gel (Sigma). Concentrations were determined using Micro BCA Protein Assay Reagent Kit (Pierce). Aliquots were kept at -80ºC, until further use.

Two new β2m constructs were made. Sea bass β2m encoding residues S1-M101

(mature protein, Fig. 4 Chapter 2) was expressed in BL21 strain after PCR incorporation of a methionine at the 5' end and a substitution of the serine-1 for an alanine (necessary for the correct NcoI restriction site), and ligation to the mentioned pET28a vector (using the NcoI/XhoI restriction sites). Sea bass β2m encoding residues M(-15)-M101 (precursor, Fig. 4 Chapter 2) was similarly expressed in BL21 strain after PCR substitution of the lysisine–(-14) for a glutamic acid (necessary for the correct NcoI restriction site), and ligation to the mentioned pET28a vector (also using the NcoI/XhoI restriction sites). Both constructs were verified by sequencing. Inductions tests were made with 0.5 mM IPTG at 37 ºC, 30 ºC and 24 ºC. Whatever the temperature, the amount of precursor β2m was very little, when compared to that of mature β2m (rβ2mS1V). Regarding rβ2mS1V, induction tests at 37 ºC resulted in higher amounts of produced protein, but all of it was insoluble. Lowering the induction temperature, reduces the total amount of produced protein, but augments its solubility. So, at 24 ºC, most of the rβ2mS1V was soluble. Soluble fraction was collected, purified through the C-terminal His tag, and buffer exchanged by dialysis (to 20 mM Tris pH 8.0/100 mM NaCl). Again, concentrations were determined using Micro BCA Protein Assay Reagent Kit (Pierce). Aliquots were kept at -80 ºC, until further use.

Human β2m recombinant protein

Human recombinant β2m was kindly provided by Prof. Sheena E. Radford (Astbury Centre for Structural Molecular Biology and School of Molecular and Cellular Biology, University of Leeds, UK). It has been over expressed in a BL21(DE3) *E. coli* strain. The protein sequence is that of human β2m with an additional N-terminal methionine, such that the polypeptide is 100 residues in length. It was diluted in dH_2O, and aliquots stored at -80 ºC until further use.

Synthetic peptides

Twelve synthetic peptides were designed by Pedro A. Reche (Department of Immunology, Faculty of Medicine, Complutense University of Madrid, Spain) based on mammalian peptide/HLA, but taking in account the sequences of the sea bass α chains, and synthesized by Sigma (Table 2). Peptides were dissolved either in 80% (v/v) dimethylformamide/dH_2O or in 50% (v/v) acetonitrile/dH_2O. Aliquots were kept at -20 ºC, until further use.

In vitro refolding

The MHC-I complex was attempted to be assembled *in vitro* from purified components, by simultaneous refolding of all constituents, by dilution of denaturing conditions at 4 ºC, with a 1:2:10 ratio of HC:β2m:peptide, according to protocol described by Garboczi *et al* [2]. A mixture of the twelve peptides was used. Controls were made in the same conditions, but not using peptides, or either using HC or β2m alone (individual refolding). After 48h in refolding buffer (400 mM L-arginine, 100 mM Tris pH 8.0, 2 mM EDTA, 0.5 mM PMSF, 5 mM reduced glutathione, 0.5 mM oxidized glutathione), samples were concentrated using Milipore stirred cells with 10 kDa cutoff polyethersulfone ultrafiltration membranes and/or 10 kDa cutoff centrifuge ultrafiltration devices (for 200 mL and 10 mL experiments, respectively). Concentrated refolded mixtures were further submitted to size exclusion chromatography and buffer exchanged (20 mM Tris pH 8.0, 150 mM NaCl, 2 mM EDTA), using either Superose 12^{TM} 10/300 GL, Superdex 75^{TM} 10/300 GL, or HiPrep 16/60 Sephacryl S-100 HR gel filtration columns (Amersham Biosciences). Collected fractions were always analyzed by SDS-PAGE, after TCA precipitation, and sometimes analyzed by mass spectrometry and/or circular dichroism. Mass spectrometry (MS) was used to evaluate the presence of bound peptide in a given fraction, and also to determine the precise molecular weight of each complex constituent, namely HC, β2m and peptide, and hence its nature. Circular dichroism (CD) was used to evaluate the secondary structure of the HCs and of β2m.

Also soluble sea bass recombinant β2m was used in new *in vitro* refolding experiments, using the previously three recombinant transcripts of HC and the same type of controls. Additional experiments were made, using human recombinant β2m (hrβ2m). As before, the MHC-I complex was attempted to be assembled *in vitro* from purified components, by simultaneous refolding of all constituents, by dilution of denaturing conditions at 4 ºC, with a 1:2:10 ratio of HC:β2m:peptide. A mixture of the twelve peptides was used. Controls were made in the same conditions, but not using peptides, or either using HC or β2m alone (individual refolding). Again, after 48h in refolding buffer (400 mM L-arginine, 100 mM Tris pH 8.0, 2 mM EDTA, 0.5 mM PMSF, 5 mM reduced glutathione, 0.5 mM oxidized glutathione), the samples were concentrated using 10 kDa cutoff centrifuge ultrafiltration devices (10 mL experiments). Concentrated refolded mixtures were further submitted to size exclusion chromatography and buffer exchanged (this time to 10 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA), using Superose 12^{TM} 10/300 GL gel filtration column (Amersham Biosciences). Collected fractions were analyzed by SDS-PAGE. Circular dichroism (CD) was used to evaluate the secondary structure of β2m.

Results

In vitro refolding with Dila-β2m (I6V-M101)

Chromatograms were identical either using or not using peptide (Fig. 6). After SDS-PAGE, it could be seen that both HC and β2m eluted separately (Fig. 6A), further confirmed by refolding experiments of HC and β2m individually (Fig. 7). Dila-UA*12 and 09 elute from the column essentially as single peaks (Fig, 7A and B), while Dila-UA*06 elutes as several smaller peaks, showing greater tendency to aggregate (Fig. 7C). β2m elutes from the column as two peaks (Fig. 7D), possibly due to dimer formation. Both concentrated refolding mixtures and eluted fractions from size exclusion of experiments with and without peptide were analysed by mass spectroscopy (MS). MS revealed a molecular weight of 33 kDa for the HC alleles and 12.2 kDa for the β2m (Fig. 8), but peptides of known molecular weight (Table 2) were not identified (Fig. 8). Although the CD plot suggests a typical curve for a protein mainly made of β strands (Fig. 9), calculated percentages of secondary structure elements suggest that refolded β2m was not in conformation with other known molecules [3-5], having ~16% α-helix, when others have 0%; and ~25% β-sheet, when others have 42-59% (Table 3). We hypothesized that this behaviour of sea bass β2m could be related to the fact that the recombinant protein starts at an equivalent position to the human recombinant β2m, based on an amino acid alignment, and not at position 1 of sea bass mature protein according to SignalP prediction. Supporting this, Esposito [6] shows that in human β2m, removal of the N-terminal hexapeptide facilitated aggregation and fibril formation. Hence, a soluble β2m without the leader peptide predicted by SignalP was produced and refolding experiments repeated.

In vitro refolding with Dila-β2m (S1V-M101)

Despite the new soluble recombinant β2m produced, chromatograms were again identical either using or not the peptide mixture (Fig. 10). In SDS-PAGE analysis, it could be seen HC and rβ2mS1V eluting together in one of the peaks. As before, Dila-UA*12 and *09 elute from the column essentially as single peaks, while *06 elutes as several smaller peaks, showing greater tendency to aggregate (not shown). Also β2m elutes from the column as two peaks. However, the first peak elutes at the same time/volume as the HC (Fig. 10), which could be shown by size exclusion of individually refolded rβ2mS1V (Fig. 11). So, although eluting together, sea bass HCs and sea bass β2m are not interacting.

In vitro refolding with human β2m

Chromatograms of refolding experiments with Dila-UA*12 and human β2m are identical either using or not the peptide mixture (Fig. 12). SDS-PAGE analysis shows that Dila-UA*12 and human β2m elute separately (Fig. 12). Differently from sea bass recombinant β2m proteins (both tested forms), human recombinant β2m elutes from the column as a single peak (Fig. 13).

Table 2

Synthetic peptides.

PEPTIDE	SEQUENCE	MW (Da)
57	QVMKNGVEL	999,18
99	CKVTHGNTV	940,06
101	VTHGNTVKY	1000,1
81	FHLTKTVPF	1071,28
102	THGNTVKYY	1064,15
21	HTSPKVQVY	1040,18
89	FTPMDGDKY	1055,17
41	TLICHVTGF	972,16
73	LAFKQDWHF	1150,34
20	KHTSPKVQV	1005,17
67	EAKQTDLAF	1004,11
79	WHFHLTKTV	1127,33

Table 3

Comparison of secondary structure elements from different β2m molecules, according to circular dicrhoism analysis.

β2m samples	α-helix (%)	β-sheet (%)	Other (%)
sea bass -15F	16.4	25.3	58.2
sea bass -5E	16.6	24.1	59.3
rβ2mS1V	19.3	13.5	67.1
human	0	59	41
chicken	0	46	52
grass carp	0	42.3	57.7

Note: human, chicken and grass carp data from references [3]; [4]; [5].

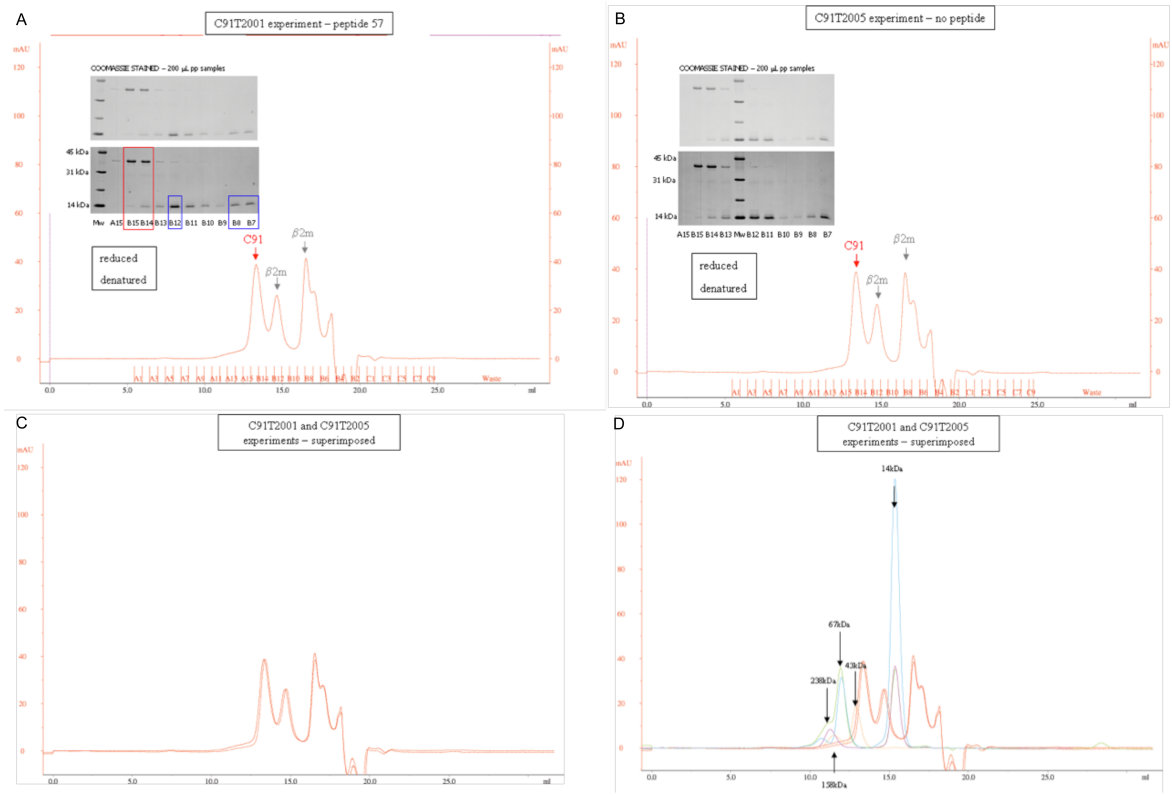


Figure 6. Refolding experiments of recombinant Dila-UA*12(C91) with denatured Dila-β2m. Chromatograms show the elution profile of an FPLC Superose 12^{TM} 10/300 GL gel filtration column. The flow rate was 0.5 mL/min, typical loading sample volumes were 0.2 mL and the elution was monitored at an absorbance of 280 nm. Experiments using recombinant Dila-UA*12(C91) and β2m with **(A)** and without **(B)** peptide are shown. Reduced SDS-PAGE gels (15%) of several fractions corresponding to the distinct peaks are also presented. At left are denoted molecular weight markers. **(C)** Superposition of the two described experiments. **(D)** Comparison of both experiments to calibration curves.

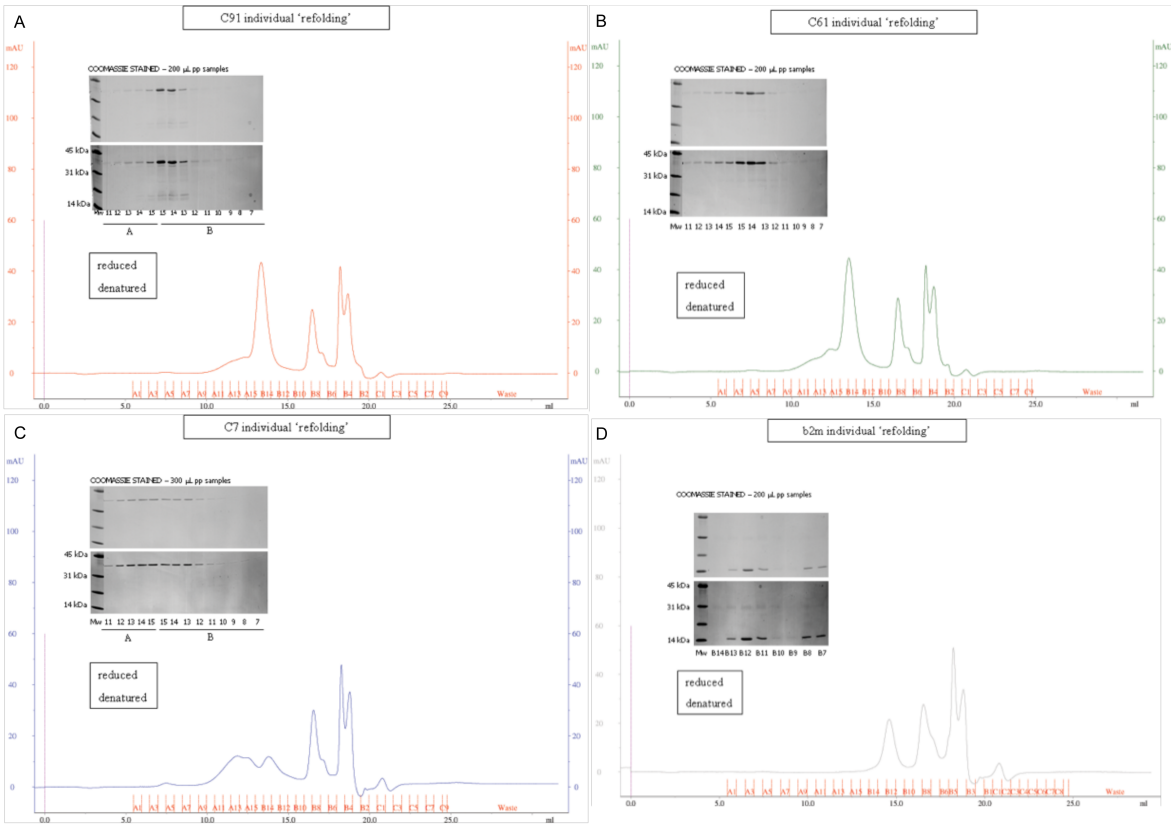


Figure 7. Refolding experiments of individual (A) recombinant Dila-UA*12(C91), (B) *09(C61), (C) *06(C7) and (D) β2m. Chromatograms show the elution profile of an FPLC Superose 12^{TM} 10/300 GL gel filtration column. The flow rate was 0.5 mL/min, typical loading sample volumes were 0.2 mL and the elution was monitored at an absorbance of 280 nm. Reduced SDS-PAGE gels (15%) of several fractions corresponding to the distinct peaks are also presented. At left are denoted molecular weight markers.

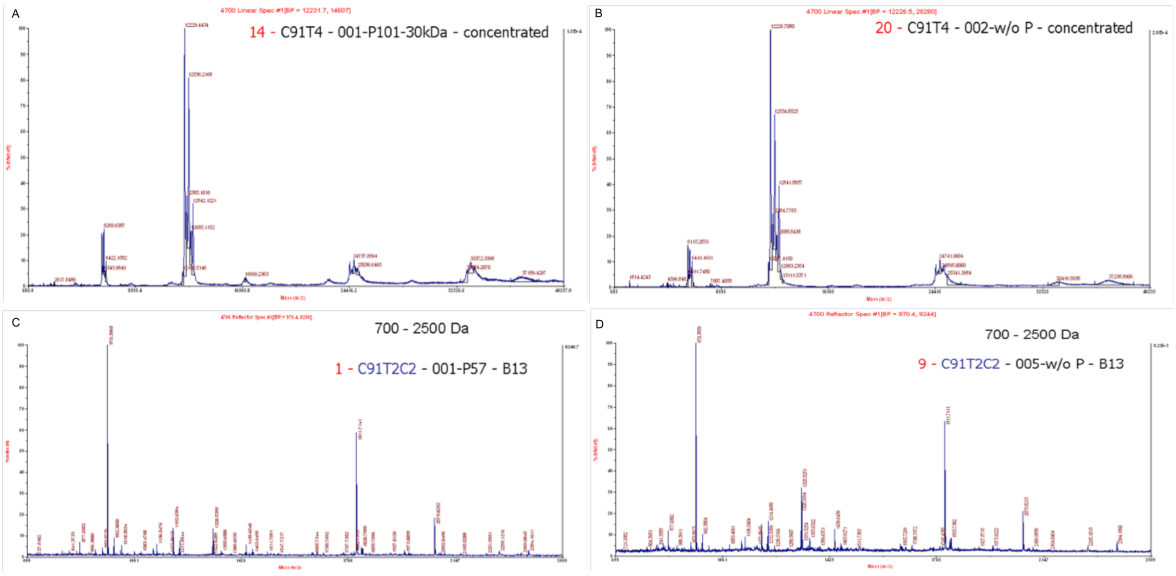


Figure 8. MS plots of recombinant Dila-UA*12(C91) and denatured Dila-β2m refolding experiments. Both concentrated refolding mixtures and fractions B13 (Fig. 8A and B) of experiments with and without peptides were analysed. The windows of mass are 0.70–40 kDa and 0.7-2.5 kDa for AB and CD, respectively.

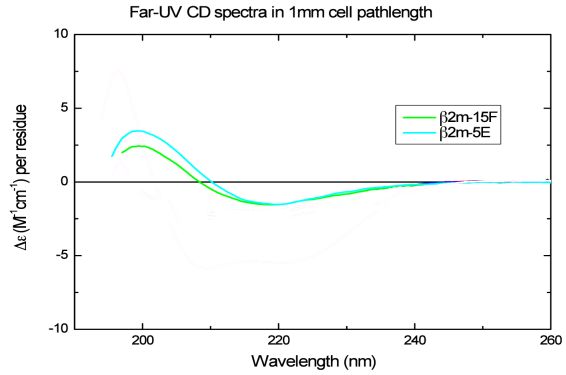


Figure 9. CD spectra from two samples of refolded recombinant Dila-β2m. The curves suggest main percentage of β strands.

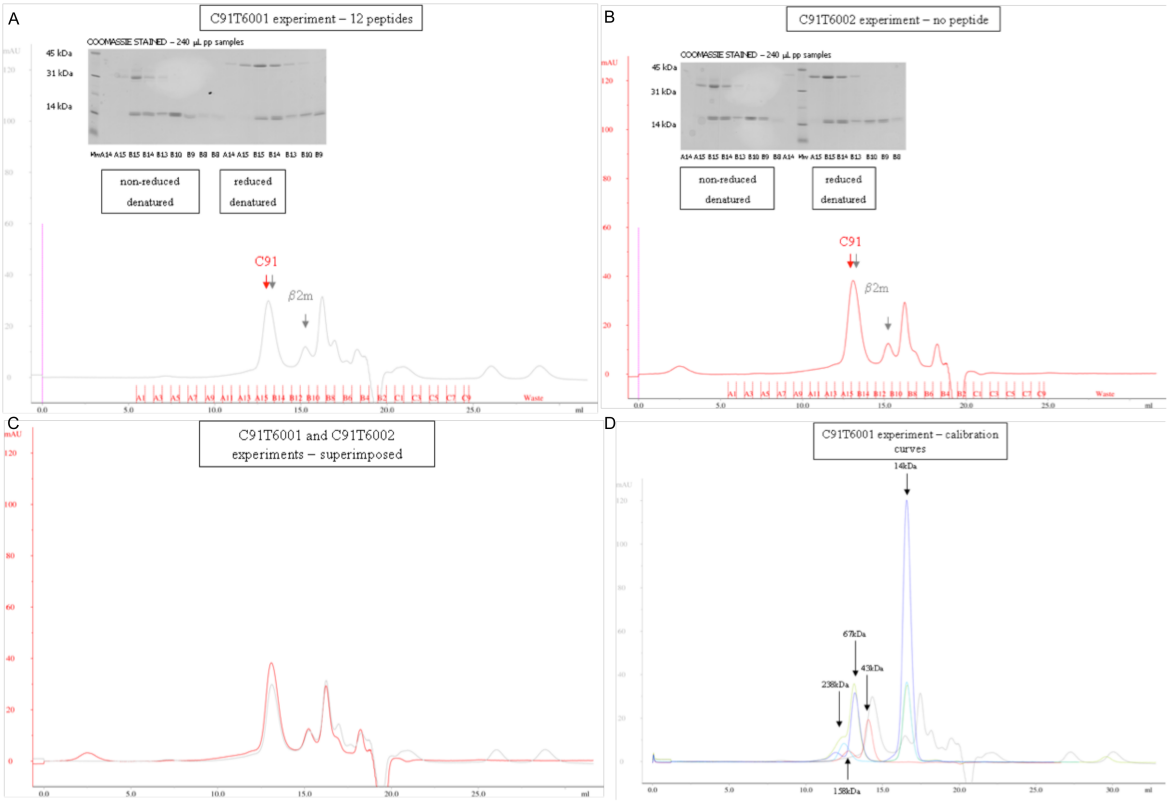


Figure 10. Refolding experiments of recombinant Dila-UA*12(C91) with soluble Dila-β2m. Chromatograms show the elution profile of an FPLC Superose 12^{TM} 10/300 GL gel filtration column. The flow rate was 0.5 mL/min, typical loading sample volumes were 0.2 mL and the elution was monitored at an absorbance of 280 nm. Experiments using recombinant Dila-UA*12(C91) and soluble β2m (rβ2mS1V) with **(a)** and without **(b)** peptide are shown. Reduced SDS-PAGE gels (15%) of several fractions corresponding to the distinct peaks are also presented. At left are denoted molecular weight markers. **(c)** Superposition of the two described experiments. **(d)** Comparison of both experiments to calibration curves.

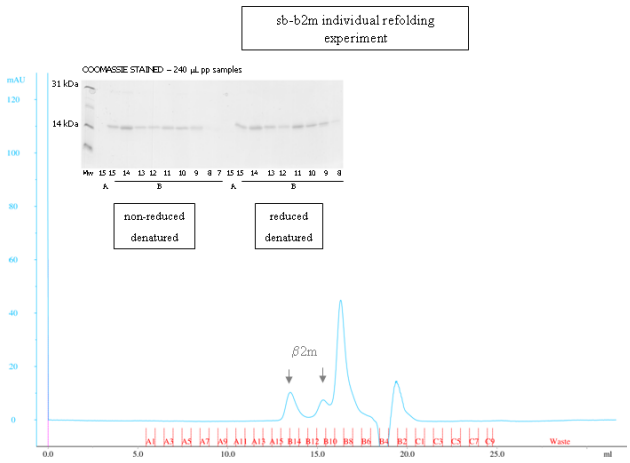


Figure 11. Refolding experiment of individual rβ2mS1V. Chromatogram shows the elution profile of an FPLC Superose 12^{TM} 10/300 GL gel filtration column. The flow rate was 0.5 mL/min, typical loading sample volumes were 0.2 mL and the elution was monitored at an absorbance of 280 nm. Reduced SDS-PAGE gel (15%) of several fractions corresponding to the distinct peaks is also presented. At left are denoted molecular weight markers.

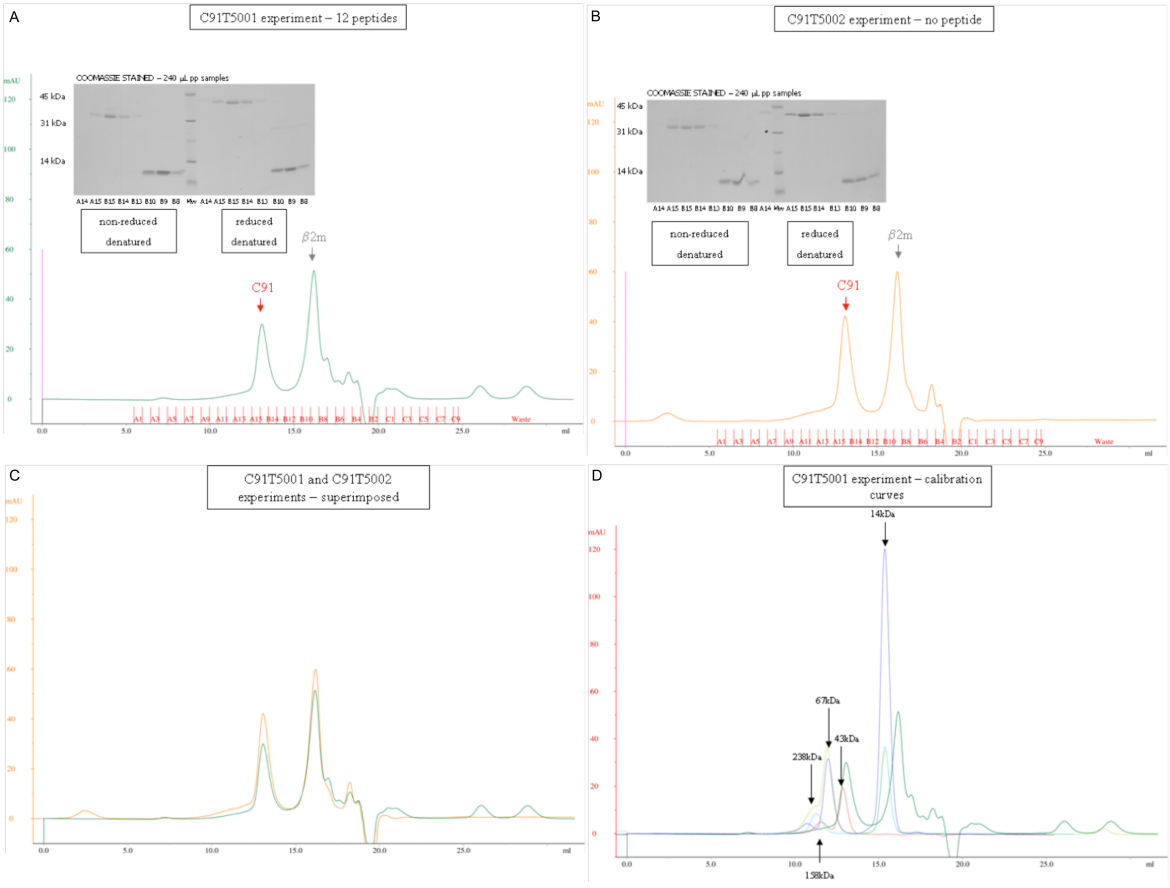


Figure 12. Refolding experiments of recombinant Dila-UA*12(C91) with human β2m. Chromatograms show the elution profile of an FPLC Superose 12^{TM} 10/300 GL gel filtration column. The flow rate was 0.5 mL/min, typical loading sample volumes were 0.2 mL and the elution was monitored at an absorbance of 280 nm. Experiments using recombinant Dila-UA*12(C91) and human β2m with **(A)** and without **(B)** peptide are shown. Reduced SDS-PAGE gels (15%) of several fractions corresponding to the distinct peaks are also presented. At left are denoted molecular weight markers. **(C)** Superposition of the two described experiments. **(D)** Comparison of both experiments to calibration curves.

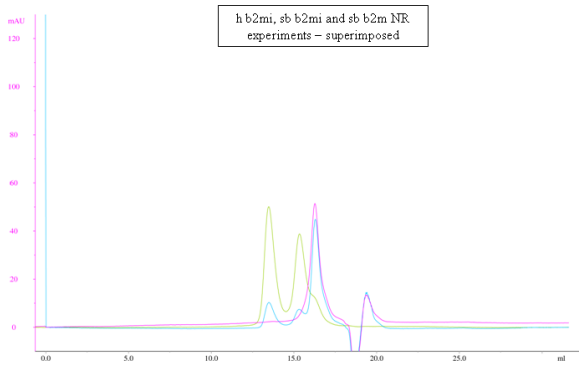


Figure 13. Comparison of chromatograms from sea bass and human recombinant β2m. Sea bass non-refolded (NR, light green) and refolded (R, light blue) rβ2mS1V experiments and human β2m individually refolding experiment (pink) are shown.

III. Transfecting a human cell line with constructs bearing sea bass (*D. labrax*, L.) recombinant MHC class I heavy chains with EGFP attached to its C terminus

As an alternative way to identify binder peptides for Dila-UA sequences, a human lymphoblastoid cell line was transfected with some of the sea bass transcripts, including representatives from both Cyt-L1 and Cyt-L2 groups. Although the transfections have worked, sea bass molecules did not reach the cell surface, precluding the initial goal. The results are summarized below.

Materials and Methods

Constructs

Plasmid EGFP-N1 was used to ligate Dila-UA*06(C7), *03(C28), *09(C61), *12(C91) or *11(C63) after PCR modification of the appropriate HC sequence to remove the stop codon, substitute lysine-(-18) for glutamic acid (for a complete correct kosak consensus sequence), and add an EcoRI restriction site at the 5' end and a SmaI site at the 3' end. All constructs were verified by sequencing; they all have an eight amino acid linker attaching HC C terminus to the N terminus of EGFP.

Transfecting 721.221 human cell line

721.221 (hereon referred to as 221 cells) is an EBV-transformed human B lymphoblastoid cell line that does not express endogenous HLA-A, -B and -C [7]. 221 cells were transfected to express different amounts of sea bass MHC class I. Cells were cultured in RPMI 1640 medium supplemented with 10% heat inactivated fetal bovine serum (FBS), 1x antibiotic and antimycotic solution, 1x sodium pyruvate, and 1x non-essential amino acids (from hereon called $RPMI^{+}$). The 221 transfectants were supplemented with 1 mg/mL geneticin (Gibco).

Generation of cells expressing different levels of sbMHC class I protein: 30 μg of the vector pEGFP-N1 either empty or containing GFP-tagged Dila-UA*06 (C7), *03(C28), *09(C61), *12(C91) or *11(C93) were linearized overnight at 37 ºC with ApaLI (Fermentas). The DNA solution was then cleaned with phenol and chloroform solutions, precipitated with 100% ethanol and washed with 70% ethanol. The DNA was resuspended in 100 μL of nuclease free water (Bioline). DNA linearization was confirmed by running the sample in 1% agarose DNA gel. For each transfection, 10^7 cells were

washed three times in RPMI 1640 medium, and the sedimented cells resuspended in 900 μL of RPMI 1640 medium. Thirty μg of DNA were added and the cells were transferred to a Gene Pulser Cuvette (Biorad), which was then placed on ice for 5 minutes. The cells were electroporated at 230 volts and 250 μF. The cells were left at room temperature for 5 min and then transferred to 50 mL tubes. Twenty four mL of $RPMI^+$ were added and the cells plated in 24-well plates (1 mL per well). After one day of incubation at 37 ºC and 7 % CO_2, 1 mL of $RPMI^+$ with 2 mg/mL geneticin was added per well. Later, 1 mL of media was removed and another 1 mL of 221 media with 1 mg/mL of geneticin was added every three days, until the media changed color. The cells were then assessed for GFP expression by flow cytometry (CellQuest™, Becton Dickinson), and positive populations with different levels of expression were sorted by dilution.

Western blot on cell lysates

Preparing cell lysates: 10^7 cells were lysed for 30 min on ice in 200 μL of lysis buffer (1% Nonidet P-40, 10 mM Tris pH 7.6, 130 mM NaCl) containing 1:100 protease inhibitors cocktail (Sigma). After centrifuging 10 min at 4 ºC, the supernatant was collected and stored at -20 ºC. These cell lysates were used for western blotting, where mouse anti-GFP mAb (clone JL-8, BD Clonetech) was used at 0.14 μg/mL and HRP-conjugated goat anti-mouse IgG Immunopure (Pierce) 16 ng/mL.

Microscopy

Cytospins were prepared with 2 x 10^5 or 5 x 10^5 cells. The appropriate volume of cell suspension was centrifuged for 5 min at 300 g and RT, washed and resuspended in 200 μL PBS. Adhesion of cells to slides was made by centrifugation for 5 min at 1000 rpm and RT. Cells were fixed with 4% paraphormaldehyde in PBS for 10 min and stopped by subsequently washing in PBS. Slides were incubated with DAPI for 15 min in the dark. Vectashield and a coverslip were added to the slides. Image analysis was performed in AxioImager Z1 (Carl Zeiss, Germany). One analysis of live cells was performed on a Laser Scanning Confocal Microscope Leica SP2 AOBS SE (Leica Microsystems, Germany).

Results

Transfection assessment by flow cytometry

Mock cells, transfected with pEGFP-N1 alone, showed GFP expression levels of 10^2 to 10^3, while the cells transfected with the different Dila-UA sequences had GFP expression levels of 10^1 to 10^2, showing, however, for each allele, a population of cells not expressing GFP (Fig. 14). In order to make these populations homogenous for GFP expression, selected transfectants were sorted by dilution. Although there was still a small population of cells not expressing GFP, the percentage of GFP-expressing cells could be improved (not shown).

Western blot

Mock cells show a band at the expected size for GFP alone (~26.9 kDa), while cells transfected with the different sea bass transcripts show either no band, one band at the size of GFP alone (~26.9 kDa), one band at the size of the recombinant (HC + GFP ~70 kDa) or even two bands one for GFP alone and one for the recombinant (Fig. 15).

Microscopy

In HC-transfectants, most samples showed no fluorescence at all (not shown); fluorescence was detected as bright spots or lines near the nuclei, especially in two samples from C91-transfected cells, suggestive of localization in compartments (Fig. 16). There was no fluorescence confined to the plasma membrane. These results show that sea bass MHC class I/GFP fusion proteins did not reach the cell surface. Perhaps they did not pass the ER quality control. We could not detect fluorescence on HC-transfectants by confocal microscopy. In mock samples (transfected with the empty vector), the cells were all fluorescent (fixed and live cells). A picture of live cells analysed by confocal microscopy is shown in Figure 16.

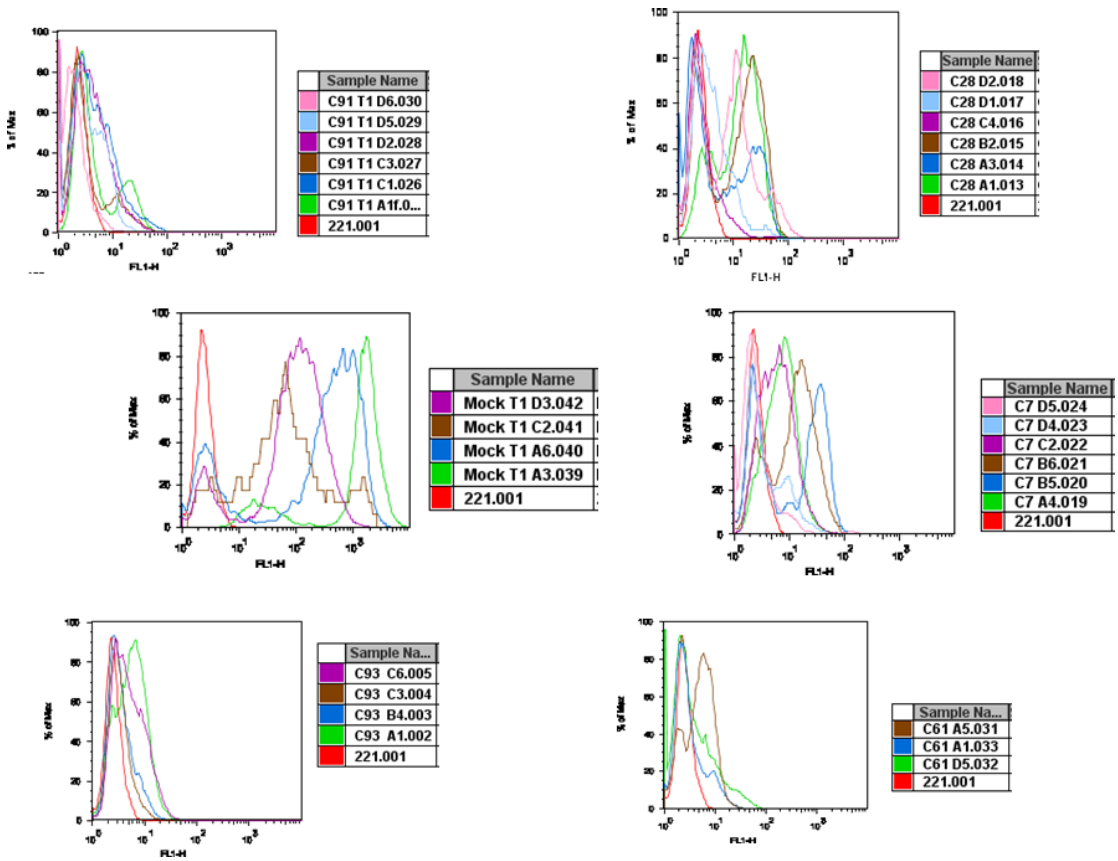


Figure 14. Flow cytometry results for GFP expression. Mock cells, transfected with pEGFP-N1 alone, showed GFP expression levels of 10^2 to 10^3, while the cells transfected with the different Dila-UA transcripts had GFP expression levels of 10^0 to 10^2, showing however for each allele a population of cells not expressing GFP. Results for each allele and for mock are presented separately. Non-transfected 221 control cells are always indicated in red.

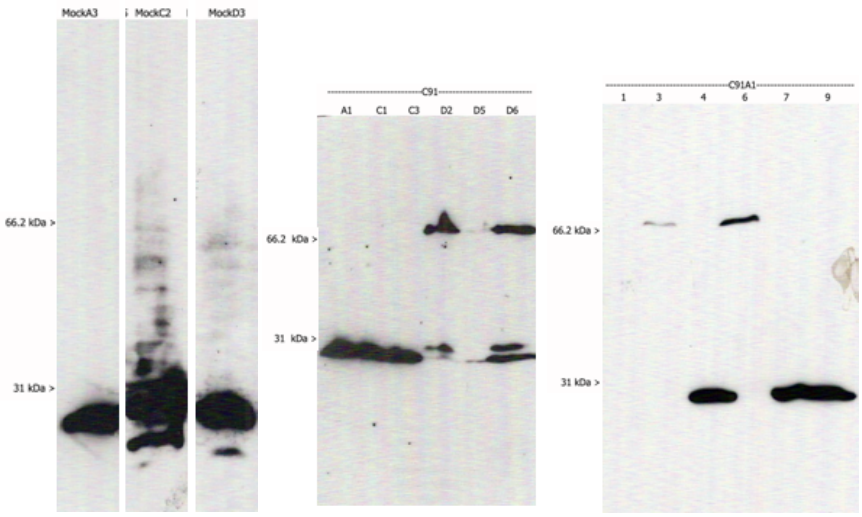


Figure 15. Western blots on transfectant cell lysates with an anti-GFP mAb. Mock cells, transfected with pEGFP-N1 alone, show a band close to the 31 kDa marker (expected size 26.9 kDa). The cells transfected with the different sbHCs show either no band, one band at the size of GFP alone (26.9 kDa), one band at the size of the construct HC+GFP (expected size 70.8 kDa for C91) or even two bands one for GFP alone and one for the recombinant HC+GFP.

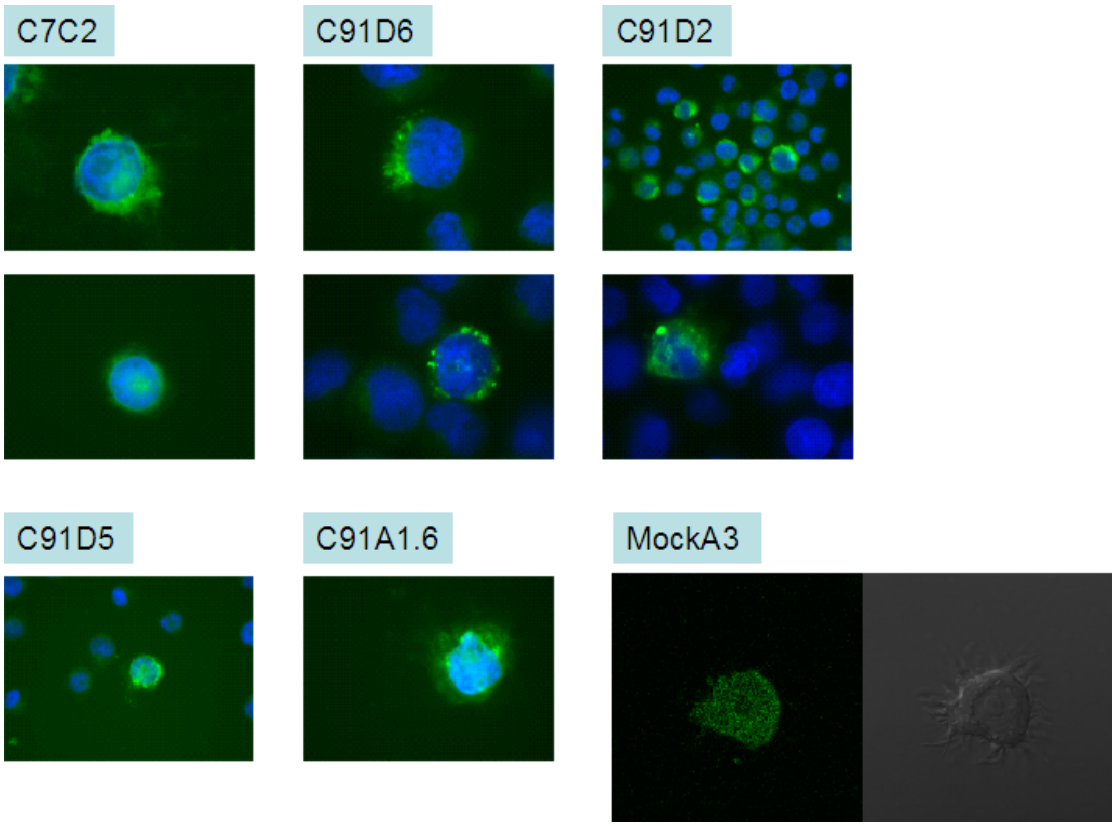


Figure 16. Pictures of fluorescence microscopy from mock and Dila-UA-transfectants where fluorescence could be detected. Dila-UA-transfectants: fluorescence was detected as bright spots or lines near the nuclei, suggestive of localization in compartments. Mock cells: all cells were all fluorescent (result on live cells).

References

1. Nagai KT, H. C. Synthesis and Sequence-Specific Proteolysis of Hybrid Proteins Produced in *Escherichia coli*. Methods in Enzymology. 1987 153:461 - 81.
2. Garboczi DN, Hung DT, Wiley DC. HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. Proc Natl Acad Sci U S A. 1992 89:3429-33.
3. Gorga JC, Dong A, Manning MC, Woody RW, Caughey WS, Strominger JL. Comparison of the secondary structures of human class I and class II major histocompatibility complex antigens by Fourier transform infrared and circular dichroism spectroscopy. Proc Natl Acad Sci U S A. 1989 86:2321-5.
4. Yan RQ, Li XS, Yang TY, Xia C. Structures and homology modeling of chicken major histocompatibility complex protein class I (BF2 and beta2m). Mol Immunol. 2006 43:1040-6.
5. Hao HF, Li XS, Gao FS, Wu WX, Xia C. Secondary structure and 3D homology modeling of grass carp (Ctenopharyngodon idellus) major histocompatibility complex class I molecules. Protein Expr Purif. 2007 51:120-5.
6. Esposito G, Michelutti R, Verdone G, Viglino P, Hernandez H, Robinson CV, et al. Removal of the N-terminal hexapeptide from human beta2-microglobulin facilitates protein aggregation and fibril formation. Protein Sci. 2000 9:831-45.
7. Shimizu Y, DeMars R. Production of human cells expressing individual transferred HLA-A,-B,-C genes using an HLA-A,-B,-C null human cell line. J Immunol. 1989 142:3320-8.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA
QREN
QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007.2013
POPH
PROGRAMA OPERACIONAL POTENCIAL HUMANO
GOVERNO DA REPÚBLICA PORTUGUESA
UNIÃO EUROPEIA
Fundo Social Europeu
SFRH/BD/42327/2007